EVALUATION OF DISPARITY MAPS

By

IVÁN MAURICIO CABEZAS TROYANO

Submitted for the Degree of Doctor of Engineering

Supervisor: María Trujillo PhD

Doctorate on Engineering with Enphasis on Computer Sciences (9702)

School of Systems Engineering and Computer Science

Universidad del Valle.

September 2013

Resumen

El problema de visión estéreo es un problema inverso y mal planteado debido a la falta de información y la inestabilidad en el sistema. Debido a lo anterior, una pequeña perturbación en la estimación de la disparidad puede causar un gran error en el cálculo de la profundidad.

Es posible encontrar en la literatura un gran número de trabajos abordando los métodos de estimación de puntos correspondientes. Por otra parte, las propuestas para evaluar cuantitativamente el comportamiento de los métodos de correspondencia estéreo sobre mediante la evaluación de mapas de disparidad son escasas. Más aun, la mayoría de dichas propuestas no tratan la estimación de disparidades como un paso intermedio en el problema de visión estéreo, desconociendo el impacto que tienen las estimaciones incorrectas en el cálculo de la profundidad. En consecuencia, podría no existir claridad en el estado-del-arte de las metodologías de evaluación, sobre cómo evaluar la precisión de los métodos de correspondencia estéreo en términos de los cálculos de profundidad.

Esta tesis reporta una investigación concerniente al juzgamiento de los métodos de correspondencia estéreo, mediante la comparación de mapas de disparidad estimados contra datos de disparidad de referencia. La pregunta de investigación formulada se describe a continuación: ¿entre un conjunto dado de métodos de correspondencia estéreo a comparar, y un determinado escenario de evaluación, cuáles son los métodos que estiman correspondencias de manera más precisa permitiendo una mejor reconstrucción de la información 3D en términos de los cálculos de profundidad?

En la tesis se presenta una metodología de evaluación para métodos de correspondencia estéreo. La metodología incluye un conjunto elementos y métodos interactuando en una secuencia ordenada de pasos. Los elementos de evaluación identificados abarcan un conjunto de imágenes de prueba, datos de disparidad de referencia y criterios de evaluación; mientras que los métodos de evaluación abarcan tanto medidas de evaluación, como modelos de evaluación. Un conjunto innovador de elementos y métodos es propuesto con el propósito de abordar la pregunta de

investigación formulada. Las contribuciones del trabajo de investigación se sintetizan a continuación:

- Se propone un fundamento teórico para los criterios de evaluación con el propósito de permitir una adecuada asociación entre el cálculo de errores y las áreas en las cuales estos se encuentran.
- Se diseñan dos medidas de evaluación que consideran tanto la magnitud del error de estimación como la relación inversa entre disparidad y profundidad.
- Se presenta una caracterización de las medidas de evaluación.
- Se presenta un modelo evaluación que aborda la comparación de métodos de correspondencia estéreo como un problema de optimización incluyendo múltiples objetivos. El modelo propuesto se basa en el concepto de dominancia de Pareto, e incluye una formulación para la interpretación de resultados.

Las propuestas son validadas en una plataforma disponible en línea, y ejemplificando su impacto sobre los resultados obtenidos mediante el proceso de evaluación, así como su relevancia con la pregunta de investigación formulada.

Palabras Clave: visión estéreo, puntos correspondientes, métodos de correspondencia estéreo, estimación de mapas de disparidad, metodologías de evaluación.

Abstract

The stereo vision problem is an inverse and ill-posed problem due to the lack of information and system instability. Thus a small perturbation in an estimated disparity may cause a large error on the calculated depth value.

In contrast to the plethora of stereo methods that have been proposed for decades, and are available in the literature, the approaches for quantitatively assessing the behaviour of stereo correspondence methods, by evaluating estimated disparity maps are not so many. Moreover, most of them do not consider disparity estimation as an intermediate step in the stereo vision problem, ignoring the impact of mismatches on depth calculations. Consequently, it may be still not clear in the state-of-the-art on evaluation methodologies, how to assess the accuracy of stereo correspondence methods in terms of depth calculations.

This thesis reports research work concerning the assessing of stereo correspondence methods, by evaluating estimated disparity maps against disparity ground-truth data. The formulated research question is as follows: Which are the method or methods accurately matching corresponding points, in order to allow a better 3D information recovery in terms of depth calculations, among a set of stereo correspondence methods being compared, under an specific evaluation scenario?.

In the thesis, an evaluation methodology for stereo correspondence methods is proposed. The methodology involves a set of elements and methods interacting in an ordered sequence of steps. Considered evaluation elements include stereo imagery test-bed, disparity ground-truth data, and evaluation criteria, whilst evaluation methods involve evaluation measures, and evaluation models. A set of innovative evaluation elements and methods are presented in order to tackle the formulated research question. In particular, the contributions of the research work can be briefly listed as follows:

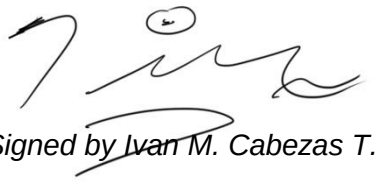
- A theoretical foundation for evaluation criteria is proposed in order to allow a proper relation between gathered errors, and challenging stereo image phenomena.

- Two evaluation measures, considering disparity estimation errors magnitude, as well as the inverse relation between depth and disparity, are devised.
- A characterisation of evaluation measures is introduced.
- An evaluation model based on the Pareto dominance relation, addressing the comparison of stereo correspondence methods as a multiobjective optimisation problem, and considering a formal interpretation of evaluation results, is proposed.

The proposals are validated using a developed on-line evaluation framework. The framework allows an interactive selection of evaluation elements and methods, exemplifying the impact of this selection, into evaluation results, as well as their relevance with the formulated research question.

Declaration

I hereby declare that I am the author of this thesis, and conducted the research to which it refers. The cited references have been consulted by me. Any idea or quotation from the work and research of a third person are fully acknowledged in accordance with the standard referencing practice of the discipline. No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of any other university or other institution of learning.



Signed by Ivan M. Cabezas T.

Acknowledgments / Agradecimientos

I would like to manifest my gratitude and sincere thanks to many people. I have a debt with some of them that I will be paying, day after day, not only with my performance but also by trying to advise and encourage many others.

To Maria Patricia Trujillo Uribe, my restless supervisor for teaching me by her example, not only about computer vision, but also on ethic and commitment for hard working, for caring about others, and for being a role model.

To Professor Panos Liatsis for sharing with me his knowledge and expertise, his advice and encouragement, as well as his sincere opinion and point of view on several topics.

To the entire staff of the Multimedia and Vision Laboratory, at the Universidad del Valle for allowing me to learn with and from them in a nice and friendly work environment.

To my father, my brother, and my sister, for their endless love and support.

To my nephews for being the ambassadors of the coming generations.

To Ruth and Antonio for caring about me as a son.

To Ruth Margaret, the other half of myself for being always supportive, enthusiastic, clever, and pragmatic. Moreover, for being my shelter, my friend, my truthful and beloved companion in this valley of shadows, laughs, happiness and tears.

I would like to dedicate this work to my mother and into her memory.

Ivan Mauricio Cabezas Troyano

Contents

Figures

Tables

Acronyms

ADC	Analog to Digital Conversion
BMP	Bad Matched Pixels
GMP	Good Matched Pixels
MRE	Mean Relative Error
SZE	Sigma– Z–Error
DOG	Difference of Gaussians
SCP	Stereo Correspondence Problem
MOP	Multi objective optimisation Problem
NCC	Normalised Cross Correlation
MSE	Mean Square Error
PSNR	Peak Signal to Noise Ratio
dB	decibels
SSD	Sum Square Differences
SAD	Sum Absolute Differences
TSAD	Truncated Sum Absolute Differences
MAD	Mean Absolute Differences
SSIM	Structural Similarity Measure

CHAPTER 1.

INTRODUCTION

Chapter Contents

1.1 Research Problem and Motivation

The stereo vision problem consists in recovering the 3D information of a scene from at least two 2D images captured at slightly different viewpoints. It is an inverse and ill-posed problem due to lack of information about depth and system instability. It has multiple applications such as: autonomous navigation (Mark & Gavrila, 2006; Ranftl et al., 2012), tele-presence (Isgro et al., 2001; Terrile & Noraky, 2012), tele-operation (Hirschmuller, 2003; Tao et al., 2011), 3DTV (Schereer et al., 2006; Dongbo et al., 2010), planetary exploration (Goldberg et al., 2002; Howard et al., 2012), and terrain analysis (Hsieh et al., 1992; Matthies et al., 2008), among others. The stereo vision problem can be tackled using the information provided by a stereo camera system, and a set of corresponding points. A pair of image points does correspond if they are projections of a same point in the 3D scene. However, the relation between corresponding pairs is not known beforehand (i.e. for a given point in the reference image, it is unknown where the corresponding point lies in the target image, and even if it really exists). Thus, the stereo vision problem entails a sub-problem: the stereo correspondence problem. The stereo correspondence problem can be addressed using a method for determining which points, in the left and in the right images, respectively,

are projections of the same 3D space point. A stereo correspondence method, as is sketched in Figure 1-1, takes as input a stereo image pair and estimates a disparity map as output. A disparity map is a computational representation of the shift between corresponding points. More precisely, disparity can be viewed as a vector, relating corresponding points. Its magnitude is inversely proportional to depth. The stereo correspondence problem involves two inherent problems: occlusion and multiple-matching. Occlusion arises when a point in one image lacks of correspondence in the conjugated image. The location of occluded points is unknown beforehand and their presence makes difficult estimating disparities of nearby image points. The multiple-matching arise due to the lack of information for uniquely identifying the correspondence of a point. It is associated to areas lacking of texture or to repetitive patterns in stereo images content.

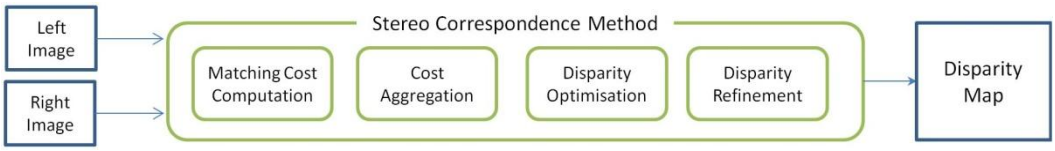


Figure 1-1 Stereo correspondence method.

The stereo correspondence problem has been widely addressed in the literature, and many methods with an increasing complexity have been proposed over the past decades. A stereo method can be described and analysed as a set of constitutive modules upon which a classification can be build. As it is shown in Figure 1-1, four main constitutive modules can be indentified (Scharstein & Szeliski, 2002): matching cost computation, cost aggregation, disparity optimisation and disparity refinement. The interaction and synergy achieved by constitutive modules is reflected on estimated disparities.

If the disparity of a point is known, the depth can be calculated by triangulation. In this way, the output of a stereo correspondence method allows to approximate a solution to the stereo vision problem. However, mismatched points may cause a large impact on the calculated depth, due to the inverse and the ill-posed nature of the problem. It gives rise to a research question:

- Which are the method or methods accurately matching corresponding points, in order to allow a better 3D information recovery in terms of depth calculations,

among a set of stereo correspondence methods being compared, under an specific evaluation scenario?.

An evaluation of estimated disparity maps can be conducted by following an evaluation methodology. It may involve multiple disparity maps, estimated for an imagery test-bed by several methods, as well as multiple evaluation criteria and measures. Although the use of an evaluation methodology is nowadays a common practice in the literature, the evaluation elements and methods commonly used present the following main drawbacks:

- Evaluation criteria are motivated and related to image phenomena that may cause innacuracies in the disparity estimation process. However, they lack of a proper formal foundation and are computationally represented as overlapping image segmentations. This overlapping implies that some image points are associated to different image pehenomena, and, consequently, if an estimation error is present in such points, it will be counted more than once during the error gatehering process. This multiple counting will cause a biasing on obtained scores. Moreover, tradittitionally used evaluation criteria do not follow the advances on state-of-the-art methods, and do not allow the evaluation on areas on which disparity asignmets are required by different application domains such as occluded areas.
- Commonly used evaluation measures are based on a counting of disparity estimation errors beyond a specified threshold. This characteristic makes the measure sensitive to the threshold selection. Moreover, such measure does not consider the magnitude on which a threshold is exceeded. In this way, a large magnitude disparity estimation error may be concealed, and considered in the same way that a small disparity error. In addition, an inherent property of disparity is ignored: its inverse relation with depth. Thus, in practice, an estimation error in a point, far from the stereo system, has a larger impact on depth calculation, that an error (of the same magnitude) in a point near to the camera system.
- Commonly used evaluation models are based on rankings. Consequently, the interpretation of obtained evaluation results may be limited, and suited more for a

contest than for a fair comparison of methods. Moreover, different conclusion may arise from the same evaluation data results, leading to a misinterpretation of the state-of-the-art.

- The selection of evaluation elements and methods used in existing methodologies is fixed beforehand. Consequently, provided evaluation scenarios are also fixed. However, in the same way that a stereo correspondence method may require tuning of parameters in order to adjust its behaviour, an evaluation process may require of different evaluation scenarios by varying the selection of elements and methods, in order to properly reveal the behaviour of stereo methods, as well as for providing useful information on which aspects a stereo method requires adjustments, and or improvements.

1.2 Investigated Approach

The investigated approach is motivated by an example. Estimated disparity maps by selected stereo correspondence methods for the Tsukuba stereo image (Scharstein & Szeliski, 2012) are shown in Figure 1-2. Figure 1-2 (a), (b), (c) and (d) illustrate disparity maps estimated by methods proposed in (Yoon & Kweon, 2005; Hosni et al. 2009; Gonzalez & Cabezas, 2009; Cabezas, 2009) respectively. The left and right views of the Tsukuba stereo image, as well as its associated disparity ground-truth data, are illustrated in Figure 1-4 (a), Figure 1-4 (b), and Figure 1-4 (c), respectively.

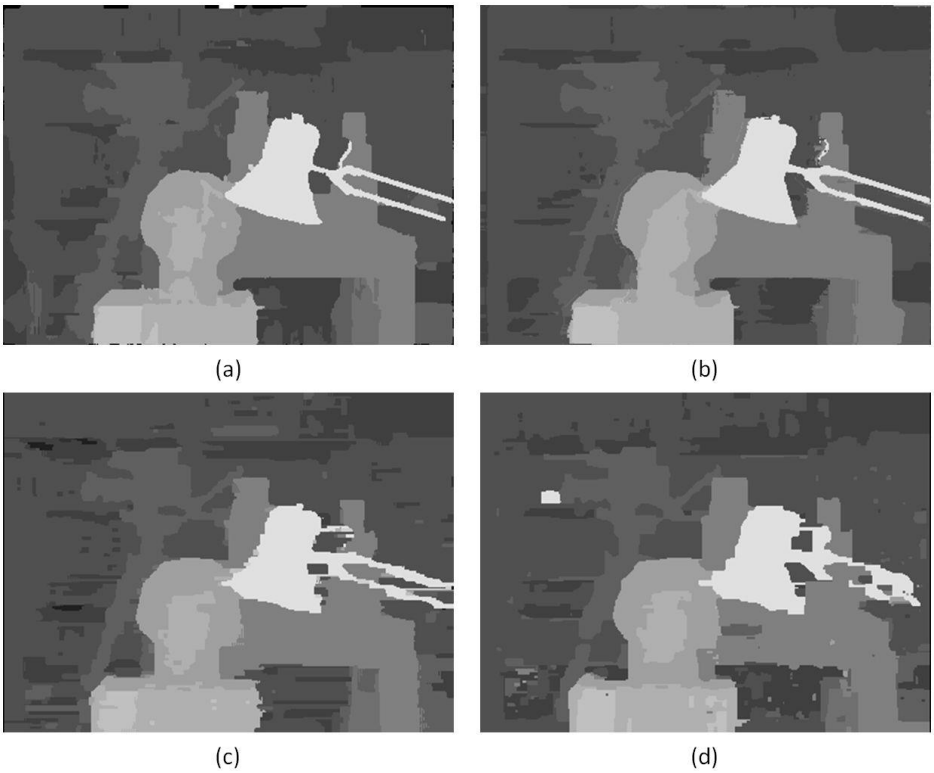


Figure 1-2 Estimated Disparity Maps for Tsukuba Image. (a) Yoon & Kweon, 2005, (b) Hosni et al. 2009, (c) Gonzalez & Cabezas, 2009, and (d) Cabezas, 2009

A visual inspection of the maps shown in Figure 1-2, by a well trained human grader allows obtaining qualitative opinions about their accuracy. A trained human grader may conclude from the map in Figure 1-2 (d) that the used method (Cabezas, 2009) has problems with foreground objects: some thin objects have disappeared, whilst others have fattened. In addition, from the map shown in Figure 1-2 (c) may be concluded that the used method (Gonzalez & Cabezas, 2009) shows poor depth discontinuities location due to the presence of some streaking artefacts. In fact, it may seem that the method used for generating the map shown in Figure 1-2 (c) is producing better results than the method used for generating the map shown in Figure 1-2 (d), but worse than the methods used to obtain the maps shown in Figure 1-2 (a) (Yoon & Kweon, 2005) and in Figure 1-2 (b) (Hosni et al. 2009). A qualitative analysis of each map is still possible when defects or artefacts in estimated maps tend to be less obvious, but the qualitative comparison among them tends to be more and more difficult (Leclercq et al., 2003). Moreover, a qualitative evaluation of disparity maps may vary largely according to several observer or observation related factors (i.e. such as observer's

experience, knowledge, fatigue, visual acuity or environmental conditions, display used, illumination, among others). In addition, the fact that a disparity map be free of noticeable artefacts or defects does not imply that the estimation is error free. In practice, a qualitative evaluation of disparity maps is a time and resources consuming task, which output is very difficult to be repeated under slightly different evaluation conditions. Moreover, obtained results by a qualitative evaluation process are inherently subjective. Hence, evaluation elements and methods interacting in an ordered sequence of steps and allow processing a large number of images in an objective, effective and efficient manner are required.

Although a quantitative evaluation process of estimated disparity maps offers advantages over a qualitative one, there are just few works addressing the quantitative evaluation of stereo correspondence methods, independently of the application domain (Guelch, 1991; Szeliski, 1999; Szeliski & Zabih; 1999 Leclerc et al., 2000; Kostliva et al., 2007), and other works within a particular domain (Hsieh et al., 1992; Mulligan et al., 2001), as well as a particular emphasis on driver assistance systems (Morales et al., 2009; Kelly et al, 2008; Morales & Klette, 2009; Klette et al., 2011; Schneider et al., 2011; Geiger et al., 2012).

On the one hand, quantitative evaluation approaches without disparity ground-truth data rely on additional views of the captured scene (Szeliski, 1999; Morales & Klette, 2009) or in a different type of ground-truth data (i.e. about the camera parameters) (Leclerc et al., 2000). However, these requirements have an impact on the generation process of stereo data (i.e. increasing the costs of the used camera system, as well as increasing the complexity of image capturing process). Moreover, these requirements are not always fulfilled in already available stereo data.

On the other hand, apart from the generation of highly reliable disparity ground-truth data, for different type of scenes (Scharstein & Szeliski 2003; Blanco et al., 2009; Smith et al., 2009; Haeuler & Klette, 2010; Geiger et al., 2012), most evaluations elements and methods commonly used still keep a resemblance with the first proposals in evaluation of disparity maps made in the early and mid 90´s by (Hsieh et al., 1992) and (Maimone & Shafer, 1996), respectively.

Nevertheless, there is still not a consensus on a standard set of evaluation elements and methods to assess stereo correspondence methods. This lack of consensus may have an impact on gauging advances on the stereo vision field, as well as on researchers and practitioners on the field. In other words, it may be not clear, among a set of stereo correspondence methods, which one, or which ones, are estimating disparity maps allowing accurate depth calculations.

The above scenario involves a decision problem:

- When is possible to determine that, comparatively, depth calculations based on a specific stereo correspondence method, are accurate, similar, or inaccurate, than depth calculations allowed by another stereo method?.

The research presented in the thesis targets the formulated research question, as a decision making problem based on multi-objective optimisation concepts and the Pareto dominance. The presented research aims to provide a formal foundation for achieving a proper quantitative evaluation of estimated disparity maps, by tackling the drawbacks of some evaluation elements and methods commonly used for the comparison of stereo correspondence methods, considering the impact of mismatches on calculated depths, and allowing a clear, unbiased and useful computation and interpretation of evaluation results.

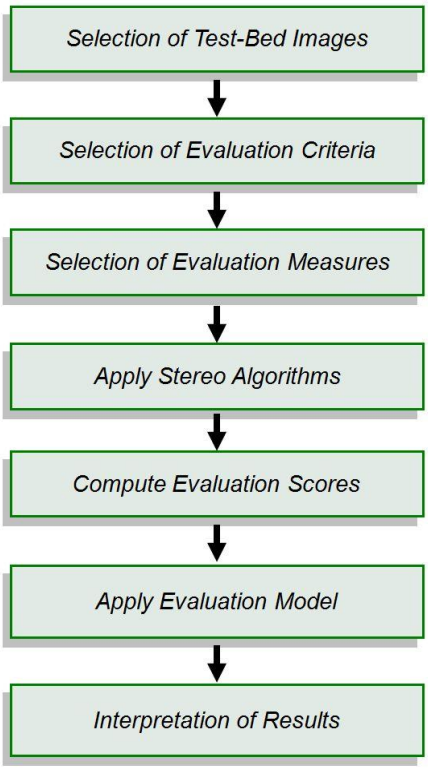


Figure 1-3 Steps involved in an evaluation methodology for stero correspondence methods.

The aim of the conducted research is motivated using the Figure 1-3, which illustrates the steps involved in a methodology for comparing stereo correspondence methods. The relevance of each one of the evaluation elements and methods used in an evaluation process is discussed. A theoretical foundation for evaluation criteria is proposed. Two evaluation measures are devised, taking into account disparity inherent properties, as well as the capabilities of a stereo camera system. A characterisation of evaluation measures is introduced. Two evaluation models based on the Pareto dominance relation are formulated, introducing a formally supported interpretation of evaluation results. Proposals are validated using an interactive on-line evaluation framework, and compared against the most commonly used evaluation elements and methods of state-of-the-art evaluation methodologies.

1.3 Data Used in the Thesis

The Middlebury's stereo benchmark data set (Scharstein & Szeliski, 2002; 2003; 2012) is used in the thesis. This data set was selected due to it is widely used and

known by the stereo vision research community. It is composed by four indoor stereo images captured under controlled conditions: the Tsukuba, the Venus, the Teddy, and the Cones, which are illustrated, with their associated disparity ground truth data, in Figure 1-4.

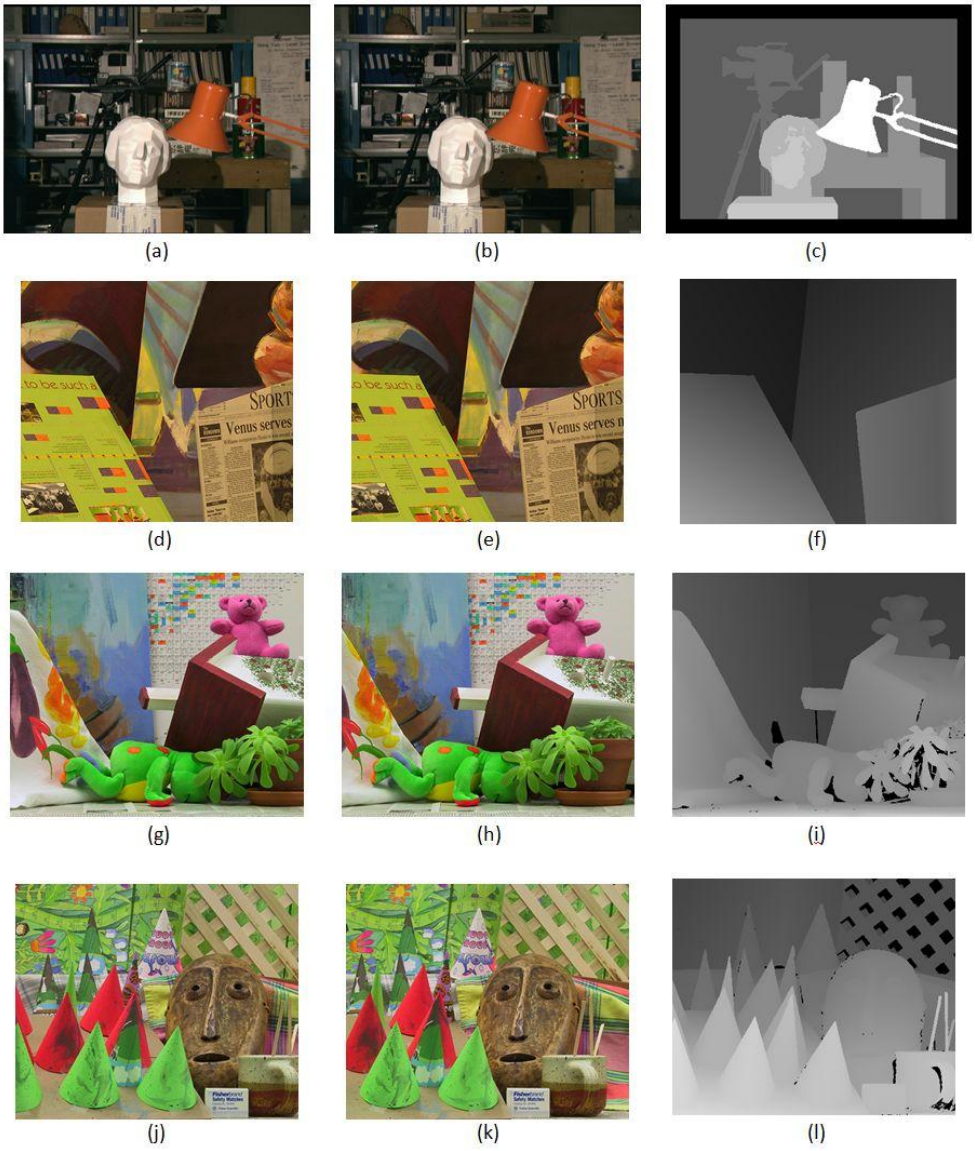
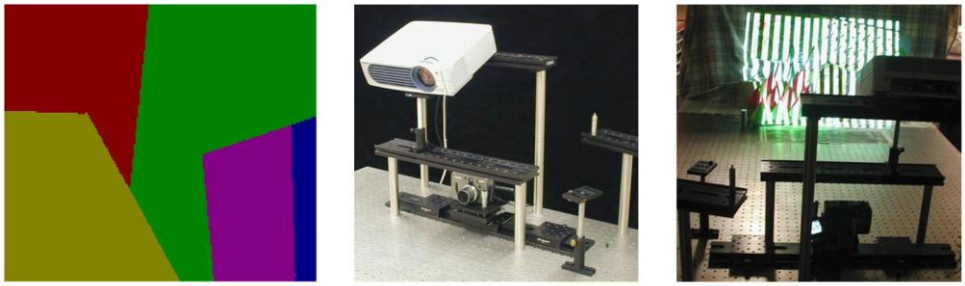


Figure 1-4 Middlebury Benchmark dataset.

The disparity ground-truth data of these images were generated by different ways. The Tsukuba stereo image is composed by front-parallel objects, and has a disparity range of 16 pixels. The disparity ground-truth data of the Tsukuba stereo image pair was generated manually and are of integer precision (Nakamura et al., 1996). It

excludes a border of 18 pixels, where no disparity value is provided. The Venus stereo image is composed by piecewise, planar slanted objects, and has a disparity range of 20 pixels. Each planar component was manually labelled, as is illustrated in Figure 1-5 (a), and a direct alignment technique (Baker & Szeliski, 1998) was used on each planar region for estimating the affine motion of each patch. The horizontal component of these motions was used to compute the ground-truth disparity map (Scharstein & Szeliski, 2002). The Teddy and the Cones stereo images contain several objects with a different geometry, and a disparity range is of 60 pixels. Their disparity ground-truth data were generated using a structured light technique (Scharstein & Szeliski, 2003). Structured light techniques rely on projecting one or more special light patterns onto a scene, usually in order to directly acquire a range map of the scene, typically using a single camera and a single projector (Boyer & Kak, 1987; Salvi et al., 2004; Koninckx & Van Gol, 2006; Chen et al., 2008; Quiang et al., 2011). The capturing setup and the illumination by structured light patterns using for the generation of the disparity ground-truth data of the Teddy and the Cones image pairs are shown in Figure 1-5 (b) and Figure 1-5 (c), respectively. In particular, a series of structured light patterns (i.e. binary Gray-code and sine waves patterns) were projected onto the scene.



(a) (b) (c)

Figure 1-5 Generation of Disparity Ground-truth data.

Projected intensities are coded to uniquely labelling each pixel. These labels are matched among the stereo views, which are illuminated from different positions. Pixels which estimated disparity do not agree under the different illumination setups are excluded from the ground-truth. Nevertheless, some areas may be shadowed under different illumination positions, and consequently the disparity ground-truth data at those regions cannot be generated. In fact, as long as each pixel is illuminated by at least one

of the projections, its correspondence in the conjugated image or even its lack of it (i.e. indicating occlusion) can be unambiguously determined.

With regard to the estimated disparity maps, inter-technique comparisons are mainly based on the maps reported to the Middlebury repository by the authors themselves. In this way, obtained results by a particular stereo method cannot be tampered by the implementation specific details missing in respective papers (Courtney et al., 1997).

1.4 Contributions

A list of the main contributions of the conducted work is presented below. Some context of each contribution is provided.

- A methodology model is presented to integrate the evaluation elements and methods in an evaluation process.
- The thesis contains a formulation of evaluation criteria based on set partitions. It allows a proper analysis of the behaviour of stereo methods regarding each evaluation criterion, without being affected by other criteria. Following the proposed formulation evaluation criteria are used in an innovative way which includes the evaluation on occluded areas.
- Two evaluation measures considering the disparity error magnitude, and overcoming the above drawbacks are presented in the thesis: Sigma-Z-Error (SZE) and Bad Matched Pixels Relative Error (BMPRE). The SZE is inherently related to the depth recovering based on estimated disparities. It does not require threshold specification by users, and is suited to be used on robotic domain applications. The BMPRE considers the error magnitude of disparity estimation over exceeding a specified threshold with regard to its inverse relation between depth and disparity. It can be used in conjunction with previously published evaluation data in order to properly quantify error impact of estimated disparity maps.

- A characterisation of evaluation measures is presented. Five criteria are identified: automatic, reliable, meaningful, unbiased and consistent. Among them, the consistent criterion is the most challenging to identify.
- An evaluation model termed $\boldsymbol{A}^*$ is proposed in the thesis. The model is based on the Pareto Dominance relation. It determines the set of stereo methods which performance can be seen as comparable (i.e. since their associated error vector values are incomparable among them in terms of Pareto dominance), and at the same time, superior to the rest of considered stereo methods. The $\boldsymbol{A}^*$ model is extended in the $\boldsymbol{A}^* - \boldsymbol{Groups}$ model, incorporating the capability of grouping the entire set of stereo methods under evaluation into groups of comparable behaviour and providing feedback for each stereo method under evaluation.
- A method for reducing the cardinality of the Pareto front is proposed in the thesis. It does not require specification of preferences nor weights by a decision maker. It is applied to the context of evaluating stereo correspondence methods to the output produced by the $\boldsymbol{A}^* - \boldsymbol{Groups}$ but it is of general purpose, and can be used during the decision making stage of any other multi objective optimisation problem.

1.5 Summary of the Candidate's Activities

The research work presented in the thesis has been reported in book chapters and conference proceedings, as is outlined below:

Book Chapters

- A Measure for Accuracy Disparity Maps Evaluation. Cabezas I., Padilla V., and Trujillo M., In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, San Martin C. and. Kim S. (Eds.), LNCS 7042, Springer-Verlag, pp. 223–231, 2011.
- A Method for Reducing the Cardinality of the Pareto Front. Cabezas I., and Trujillo M., In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Alvarez L. et al., (Eds.), LNCS 7441, Springer-Verlag, pp 829-836, 2012.

- Methodologies for Evaluating Disparity Estimation Algorithms. Cabezas I., and Trujillo M., In: Robotic Vision: Technologies for Machine Learning and Vision Applications, García-Rodriguez J. and Cazorla Quevedo M. (Eds.), IGI Global, 2013.

Conference Proceedings

- A Non-Linear Quantitative Evaluation Approach for Disparity Estimation - Pareto Dominance Applied in Stereo Vision. Cabezas I., and Trujillo M., In: Proceedings of the International Conference on Computer Vision Theory and Applications - VISAPP, SciTePress, pp. 704-709, 2011.
- An Evaluation Methodology for Stereo Correspondence Algorithms. Cabezas I., Trujillo M. and Florian M., In: Proceedings of the International Conference on Computer Vision Theory and Applications - VISAPP, SciTePress, pp. 154-163, 2012.
- On the Impact of the Error Measure Selection in Evaluating Disparity Maps. Cabezas I., Padilla V., Trujillo M., and Florian M., In: Proceedings of the IEEE World Automation Congress - WAC, pp. 24-28, 2012.
- BMPRE: An Error Measure for Evaluating Disparity Maps. Cabezas I., Padilla V., and Trujillo M., In Proceedings of the IEEE International Conference on Signal Processing - ICSP, vol 2, pp. 1051-1055, 2012.

1.6 Thesis Outline

The thesis comprises six chapters. It is structured and organised as follows:

- An overview of different concepts and required background for a better understanding of the topics discussed in the thesis is provided in Chapter 2. It begins by considering the components of a mathematical model of imaging, describes the stereo correspondence problem and their two inherently associated problems, some of the constraints that make possible addressing the stereo correspondence problem using a computer, as well as the stratified reconstruction approach. Chapter 2 ends pointing out most commonly used distance functions and presenting the basics of multi-objective optimisation.

- A review and a discussion on related approaches as well as on stereo correspondence methods are presented in Chapter 3. It includes the existing classifications of stereo methods, stereo methods used for matching key points in unrectified stereo image pairs, stereo methods for matching points in 1D after the imposition of the epipolar constraint doing special emphasis on local methods, modular components used in post-processing stages, as well as evaluation methodologies with and without disparity ground-truth data. Chapter 3 ends with a brief review on decision making strategies in multi objective optimisation.
- The proposals of the thesis are presented, formulated and discussed in Chapter 4. The evaluation element and methods, as well as the sequence of steps followed in an evaluation methodology of stereo correspondence methods are explicitly identified. A formulation for evaluation criteria based on sets partition is presented. Two new measures for comparing estimated disparity maps against disparity ground-truth data are proposed. Their differences and advantages over conventionally used evaluation measures are discussed. A set of criteria for selecting an evaluation measure during the evaluation process is proposed and discussed. A method for reducing the cardinality of the Pareto Front in the context of multi objective optimisation decision making is presented. Chapter 4 ends with a discussion of a proposed adaptive online evaluation framework for the comparison of stereo correspondence methods.
- The experimental evaluation and the validation of the proposals are presented in Chapter 5. This Chapter is devoted to the discussion of the impacts of presented proposals. The experimentation and the validation of proposals are performed over the interactive and online developed evaluation framework.
- Finally, Chapter 6 concludes the thesis by summarising the research outcomes, highlights the contributions of this work, as well as pointing out work for future research in the addressed domain.

CHAPTER 2.

THEORETICAL BACKGROUND

2.1 Image Formation Process

The first step in the vision process is image formation. It occurs when a sensor registers irradiance interacting with (i.e. being reflected or radiated by) physical objects. (Ballard & Brown, 1982). A mathematical model of imaging may involve, among others, the following components:

- An intensity function, it is the fundamental abstraction of an image.
- A geometrical model, it is used to represent how the 3D world is projected into 2D images.
- A radiometric model, it is built to represent how light sources and reflectance properties may affect irradiance measurement at the sensor.
- A digitising model, it describes the process of obtaining discrete samples.

A digital image can be described as a matrix of irradiances or intensity discrete samples, stored in a computer using a previously fixed limited precision. A definition of monochromatic (i.e. grey-level) digital image is given as follows:

Definition 2.1: Let b be a quantity of bits previously fixed. Let i be a function, $i: (n \times m) \rightarrow p$, such that n, m and $p \in \mathbb{N}$, subject to: $1 < n \leq N, 1 < m \leq M, 0 \leq p \leq P$, where $P = 2^b - 1$, and N and M are the amount of rows and columns in a grid accommodation, respectively.

In the above definition, a p value is an intensity value in a picture element or a pixel, whilst n and m values are spatial coordinates. In this way, an intensity function relates a pair of spatial coordinates to a pixel value. In practice, a pixel value is not determined by a single point in 3D space, but by a small area in a surface.

2.1.1 Geometrical Models in Imagery

The relation between the 3D world and digital images can be modelled by geometry. The following notation is used for presenting some basic concepts of imaging geometry:

- $(X_w, Y_w, Z_w)^T$, a 3D point in world coordinate system.
- $(X, Y, Z)^T$, a 3D point in camera coordinate system.
- f , focal length (the distance between the image plane and the optical centre).
- $(x, y, f)^T$, a 2D point in image coordinate system.
- $(u, v, f)^T$, a 2D point in pixel coordinate system.
- $(x, y, 1)^T$, a 2D point in normalised image coordinate system.
- $(u, v, 1)^T$, a 2D point in normalised pixel coordinate system.
- E , the 3x3 essential matrix.
- F , the 3x3 fundamental matrix.

A camera is, in brief, an electronic sensing device capable of producing a mapping between the 3D world and a 2D image. In this mapping, there are inherent errors such as quantisation effects or geometric distortions, among others. From a geometric point of view, a camera can be analysed as a perspective projection model. In this regard, the Pinhole camera model is an ideal model.

2.1.1.1 Pinhole Camera Model

A pinhole camera can be constructed by an enclosure or a box with a small hole in the front of it. An inverted image is formed on the back of the box, when rays coming from world objects passing through the pinhole. A pinhole camera is illustrated in Figure 2-1 (Forsyth & Ponce, 2011).

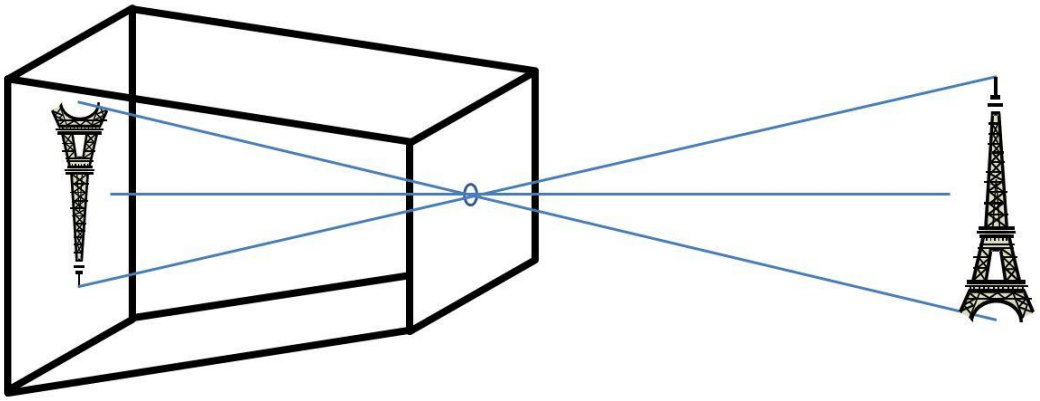


Figure 2-1 Pinhole camera.

In this construction, the front of the box is acting as the focal plane. The pinhole, – denoted as C –, is the optical centre. The back of the box – denoted as I – is the image plane. A pinhole camera model is a mathematical idealisation of the construction described above, on which a virtual image plane I is in front of the focal plane, at a distance f, obtaining a non-inverted image. The pinhole camera model, with a virtual image plane, is illustrated in Figure 2-2. In this model, light rays are converging at the optical centre C. The optical axis is the line going through the optical centre and perpendicular to the image plane. It is intersecting with the image plane at the principal point, – denoted as c.

In order to establish a relation between, 3D camera coordinates, and 2D image coordinates, it is necessary to define coordinate systems. The image coordinate system (c, x, y) is defined with the origin at c, and its axes are determined by the camera scanning and sampling system. The 3D camera coordinate system (C, X, Y, Z) is defined with the origin at C.

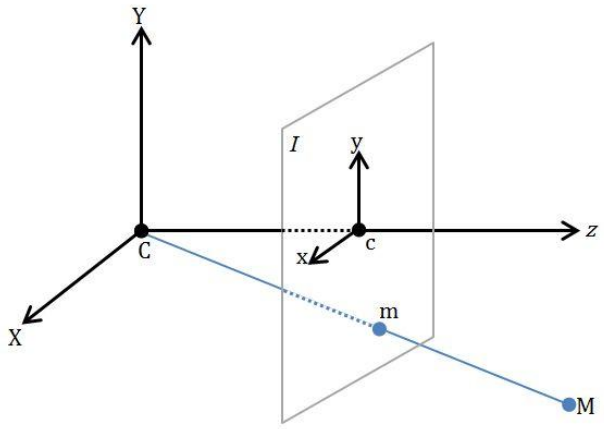


Figure 2-2 Pinhole camera model.

The Z-axis coincides the optical axis of the camera, and the X and Y axes are parallel to the image coordinate system x- and y- axes. In addition, the relation between the 3D camera coordinate system and the image coordinate system is defined as follows:

$$\frac{x}{X} = \frac{y}{Y} = \frac{f}{Z} \tag{2.1}$$

Thus, a 3D point $(X, Y, Z)^T$ is expressed in the image coordinate system according to:

$$x = \frac{X f}{Z}, \tag{2.2}$$

and

$$y = \frac{Y f}{Z}. \tag{2.3}$$

The pinhole camera has two main weaknesses that make impractical its use (Xu & Zhang, 1996). First, an ideal pinhole, of an infinitesimal aperture does not allow the gathering of enough amount of light in order to measure brightness. Second, the diffraction that occurs at the pinhole, in addition with the wave nature of light, implies that a larger fraction of the incoming light is deflected far from the direction of the incoming ray. These weaknesses may be alleviated by a lenses system.

2.1.1.2 Thin Lens Model

A thin lens is an ideal lens, which should have the same smooth curvature in booth sides. By construction, a thin lens deflects all rays parallel to the optical axis and coming from one side onto the focus of the other, following two properties:

- A ray passing through the centre of the lens is undeflected.
- All rays parallel to the optical axis converge to a point, on the other side of the lens, at a distance equal to the focal length.

A thin lens model produces the same projection as the pinhole model, but gathering enough amount of light. A cross-sectional view of a thin lens sliced by a plane containing the optical axis is illustrated in Figure 2-3.

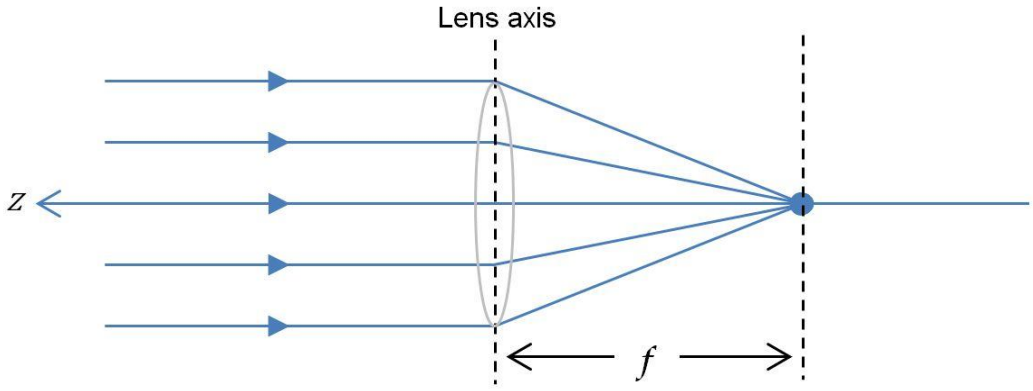


Figure 2-3 - Light convergence in a thin lens camera model.

In addition, a thin lens model involves two concepts: the depth of field and the field of view. The depth of field is a range of distance along the optical axis on which objects are properly focused. Objects outside of this range will appear blurry in the image. The field of view is an angular measure of the portion of 3D space which is captured by a camera.

2.1.1.3 Distortion Enhanced Models

The final coordinates on the image plane of a 3D world point may be affected due to several types of imperfections in the design and assembly processes of the lens. In fact the magnitude of the geometric distortions depends on the quality of the used lens. Geometric distortions are related to the curvature of the lens, and cause a

displacement of a given point, from its ideal location. Thus, a straight line in the 3D world is not projected onto a straight line in the image plane. The radial distortion is the most common distortion (Heikkilä, 2000; Zhang 2000). Radial distortions are symmetric about the optical axis. There are two types of radial distortions: barrel and pincushion. A barrel distortion is related to a negative radial displacement, causing that outer points be grouped together and a decreased scale. A pincushion distortion is related to a positive radial displacement. It causes outer points be spread and an increased scale. In this work distortion is not considered.

2.1.1.4 Camera Parameters

A reconstruction of the 3D structure of a scene requires a link between the coordinates of a point in 3D space, with the coordinates of their projection onto the image plane. A 3D point M in space is, regarding to the world reference frame, located in the coordinates $(X_w, Y_w, Z_w)^T$. It is, with regard to the camera reference frame, located in coordinates $(X, Y, Z)^T$. The point $M$is projected onto the image point m, which is the intersection of the line joining M and (the optical centre) C, with (the image plane) I. The image point m has coordinates $(x, y, f)^T$. In fact, the third component of an image point is always equal to the focal length. Thus, as we saw in section 2.1.1.1, there is a mapping from Euclidean 3-space, to Euclidean 2-space, as follows:

$$(X_w, Y_w, Z_w)^T \rightarrow \left(\frac{X\,f}{Z}, \frac{Y\,f}{Z}\right) \tag{2.4}$$

It is often assumed that the camera reference frame can be located with respect to some other, known, reference frame (i.e. the world reference frame), and that the coordinates of image points in the camera reference can be obtained from pixel coordinates, which are available from the image. This is equivalent to assuming knowledge of some camera characteristics. These camera characteristics are known as the camera intrinsic and extrinsic parameters.

2.1.1.5 Extrinsic Parameters

The extrinsic parameters are a set of geometric parameters that uniquely identify the transformation between the unknown camera reference frame and a known world reference frame. Extrinsic parameters can be expressed in terms of a 3D rotation

followed by a translation, as is illustrated in Figure 2-4. The rotation is based on a 3x3 orthogonal matrix R, and the translation is based on a vector t. Thus, a mapping between the world reference frame and the camera reference frame is defined as follows:

$$(X, Y, Z)^T = R\,(X_w, Y_w, Z_w)^T + t. \tag{2.5}$$

The matrix R can be parameterized by the Euler angles yaw, pitch and roll. These angles, in conjunction with the tree components of the vector t, are the components of the extrinsic parameters.

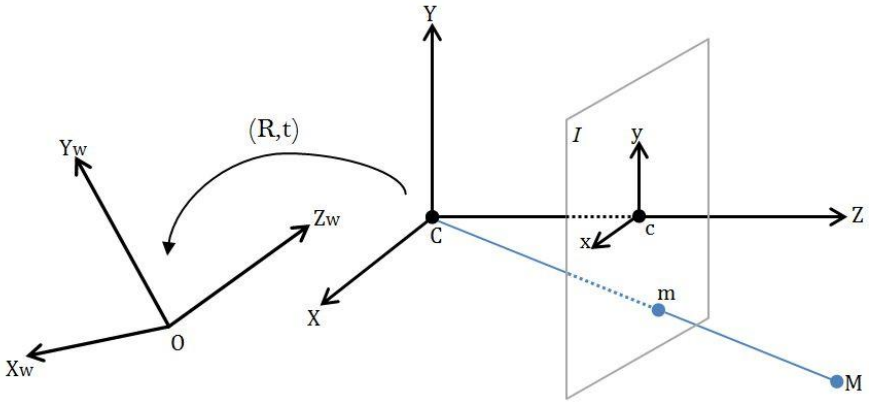


Figure 2-4 Relation between the camera and the world coordinate system.

2.1.1.6 Intrinsic Parameters

The intrinsic parameters are the factors that define the scanning and the sampling relation between the image coordinate system and pixel coordinate system. They characterise inherent optical, geometric and digital properties of a camera.

Intrinsic parameters involve the perspective projection, the transformation between the camera frame coordinates and pixel coordinates, and the geometric distortion introduced by optics. The focal length is the only parameter related with the perspective projection. The pixel coordinate system (o, u, v) is used in the digitised image. Let (u_0, v_0) be the coordinates of the principal point c. Let k_u and k_v be the horizontal and vertical scale factors, respectively, whose inverse characterise the size of pixel in word coordinate units. Then, neglecting any geometric distortions introduced by

the optics, the transformation between camera frame coordinates and pixel coordinates are expressed by the coordinates in pixels of the image centre(u_0, v_0) , or principal point, and the effective size of the pixel in the vertical and horizontal directions, (k_u , k_v), respectively, as follows:

$$x = \frac{u - u_0}{k_u}, \tag{2.6}$$

$$y = \frac{v - v_0}{k_v}. \tag{2.7}$$

The relation between the image and the pixel coordinate system is illustrated in Figure 2-5.

With regard to radial distortions, they are commonly ignored when high accuracy is not required, since they are usually very small.

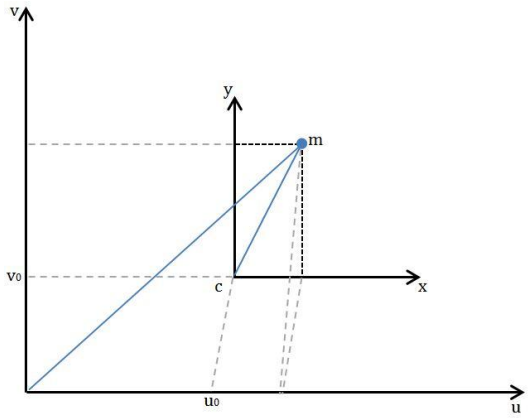


Figure 2-5 Relation between the camera and the world coordinate system.

2.1.2 Radiometric Models

Radiometry is the part of image formation process concerned with the relation among the amount of light energy emitted from light sources, reflected from surfaces, and measured by sensors (Trucco & Verri, 1998). Light sources can be divided in two main categories: point sources (i.e. the sun lighting at infinity) and area sources (i.e. a fluorescent lamp in lab ceiling). The light hitting an object is scattered and reflected (Szeliski, 2010). Thus, two radiometric issues have to be considered during the image formation process: how much of the illuminating light is radiated (i.e. emitted or reflected)

by objects surface (i.e. scene radiance), and, how much of the reflected light effectively reaches the image plane of a camera (i.e. image irradiance).

With regard to scene radiance, a surface reflectance model specifies how a surface reflects light. The Lambertian surface reflectance model assumes that objects' surface appears equally bright from all viewing directions. It applies, for instance, to nonspecular surfaces, as well as for several materials and finishes such as a matte paint, or a piece of paper, among others. Consequently, there are several circumstances on which surface reflectance do not follow the Lambertian model.

With regard to image irradiance, it can be assumed, in general terms, as uniformly proportional to the scene radiance over the whole image plane.

2.1.3 Digitising Models

After the light is reflected by object surface, and passes through camera's optics, it reaches the imaging sensor (Szeliski, 2010). This light energy (i.e. incoming photons) is then converted to electrons by the sensor. The two main kinds of sensors used in cameras are the Charged Coupled Device (CCD) and the Complementary Metal Oxide on Silicon (CMOS). Each one of these sensor perform their task using a variety of technologies, with advantages and drawbacks of their own.

In addition, there are several factors affecting the image sensor performance such as such as: exposure time, sampling, fill factor, analog gain, ADC resolution, and digital post-processing, among others (Szeliski, 2010).

- The exposure time is determined by the shutter speed. It controls the amount of light reaching the image sensor. An improper amount of light may turn into an over or an under exposed image.
- The sampling pitch is the distance among neighbouring sensor cells on the image chip. A sensor with a smaller sampling pitch provides a higher resolution, but, at the same time, a smaller pitch implies that each sensor has a smaller area making it less light sensitive and prone to noise.
- The fill factor is related to the active sensing area size. A higher fill factor is capable of capturing more light, but it also requires more electronics.

- A higher analog gain allows the camera to perform better under low light conditions.
- The Analog to Digital Conversion (ADC) is the last step occurring within an imaging sensor. It has two aspects of main interest, the resolution, and its noise level. The resolution is related to the quantity of bits involved, whilst the noise level is related to how many of these bits are useful in practice.

Digital post-processing is related to a set of operations performed by a camera after the conversion of irradiance values to digital bits, and previous to the compression and storing of pixel values. They may include luminance mapping and colour demosaicing, among others.

2.2 Stereo Correspondence

In a stereo camera system, a 3D scene – the input – is simultaneously captured from slightly different points of view, and a stereo image is produced as the output, according to the projective model of the system. Thus, the point M in 3D space may be captured in both of the views generated by the stereo system. These two conjugated projections of the 3D point are corresponding points. A definition of stereo correspondence is provided as follows:

Definition 2.2: Let M be a 3D space point in the world coordinates. Let $\mathrm{m}_l = (\mathrm{x}_l, \mathrm{y}_l, 1)^T$ be a point on the left image plane, in normalised image coordinates. Let $\mathrm{m}_r = (\mathrm{x}_r, \mathrm{y}_r, 1)^T$ be a point on the right image plane, in normalised image coordinates. The points m_l and m_r are corresponding points, if and only if they are projections of the point M.

An inverse problem arises when the goal is to recover the 3D structure of a scene from a stereo image. A 3D information recovering process is possible if the information about the correspondences is known and some information about the stereo camera system is available. In this case the output of the system is given, and the original input can be estimated.

2.2.1 Stereo Correspondence Problem

The stereo vision process, defined as the recovery of the 3D structure of a scene from 2D images, is based on the information about corresponding points. However, in practice, the information about stereo corresponding points is unknown beforehand. Thus, the stereo correspondence problem is associated to a lack of knowledge about the existence and the location of the underlying correspondences that exist in a stereo image. There are two inherent problems to the stereo correspondence problem: occlusion and multiple matching. The occlusion problem arises when only a single projection of a 3D point is captured into the stereo image pair. In this case, the original depth of such point cannot be recovered from the stereo image. The multiple matching problem arises when the image content in the stereo pair does not uniquely identify which are the conjugated matching points. The multiple matching problem is associated to areas lacking of texture, as well as to the presence of repetitive patterns, among others.

The stereo correspondence problem can defined as follows.

Definition 2.3: Let S be a stereo camera system. Let $I_l$be the left image plane of S. Let I_r be the right image plane of S. Let M be a 3D space point in world coordinates. Let $\mathrm{m}_l = (\mathrm{x}_l, \mathrm{y}_l, 1)$ be the projection of M onto I_l, in normalised image coordinates of the left camera reference frame. Let $\mathrm{m}_r = (\mathrm{x}_r, \mathrm{y}_r, 1)$ be the projection of M onto I_r, in normalised image coordinates. The stereo correspondence problem has the following characteristics:

- It is not known beforehand if the point m_r really exists.
- If the point m_r does exist, the values of their image coordinates are unknown.

The stereo correspondence problem is an ill-posed problem due to the lack of information about depth and the instability of the solution of the system. As a consequence of instability, a small perturbation in the matching of conjugated points may produce a large error in the 3D information recovery process.

2.2.2 Disparity Estimation

The stereo correspondence problem can be addressed as a search problem guided by an optimisation strategy: for each point in the reference image, its matching point is searched in the target image, according to the scores computed by matching functions. Commonly, the left view is used as the reference image, and the right view is used as the target image. However these roles are interchangeable.

The vector relating corresponding points is termed disparity. A disparity vector can be defined as follows.

Definition 2.4: Let $m_l = (u_l, v_l, 1)$ be a point in the left image plane I_l, in normalised pixel coordinates. Let $m_r = (u_r, v_r, 1)$ be a point in the right image plane I_r, in normalised pixel coordinates. Let m_l and m_r be corresponding points. Let $\vec{d}$ be the disparity vector relating points m_l and m_r, $\vec{d} = (u_l - u_r, v_l - v_r)$.

Once a correspondence has been established, measuring its associated disparity is straight forward. Thus, errors in disparity assignments are due to inaccurate or wrong matches. In general terms, there are three types of errors during the search for matching points: mismatches, false negatives and false positives (Goulermas, 2000). A mismatch occurs when a point in the reference image is matched with the wrong point in the target image. A false negative occurs when a point which should be matched is left unmatched. A false positive occurs when a point which should be left unmatched is given a match.

The magnitude of a disparity vector is inversely proportional to the depth. A larger value indicates a closer distance between a point and the camera system. In the context of stereo vision, the set of disparity vectors for a particular stereo image is termed as disparity map. It can be represented as an image, being the points of the reference image, the spatial domain.

2.2.3 Stereo Constraints

In a strict mathematical sense, an ill-posed problem cannot be solved. An approximation to the solution can be achieved based on *a priori* information and imposing constraints. The constraints that are commonly considered by different approaches to the stereo correspondence problem are presented below.

2.2.3.1 Epipolar Constraint

In a stereo camera system, there arises a relation called the epipolar geometry. The epipolar geometry is independent of the scene structure (Faugeras, 1993; Xu & Zhang 1996). It is explained as follows:

Let C_l and C_r be the optical centres of the left and right cameras, respectively, as it is shown in Figure 2-6.

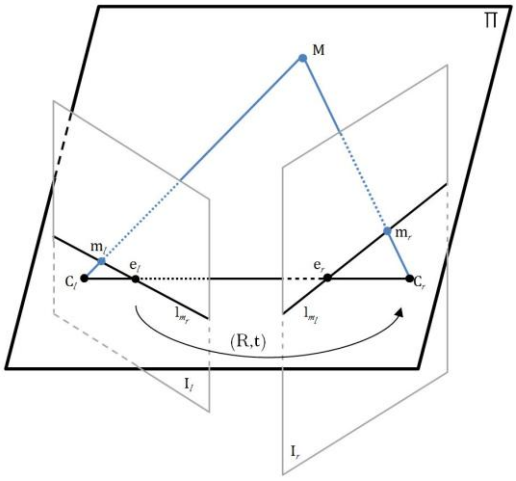


Figure 2-6 Epipolar constraint in a convergent camera model.

Given a point m_l in the I_l image plane, its corresponding point m_r in the I_r image plane is constrained to lie on a line termed the epipolar line of m_l, denoted by l_{m_l}. The line l_{m_l} is the intersection of the plane Π, defined by m_l, C_l and C_r, with the I_r image plane. This plane is the epipolar plane. This is due to the image point m_l correspond to an arbitrary point on the segment $(\overline{C_l\ M})$, and the projection of such segment on I_r is the line l_{m_l}. The line $(\overline{\mathrm{C}_l\ \mathrm{C}_r})$ is called the baseline. The baseline intersects the two image planes I_l and I_r at the points, e_l and e_r respectively. These points are called epipoles. The epipolar lines in image plains $I_l$and I_r, are in intersection with the epipoles e_l and e_r, respectively. There is symmetry in the restriction of where the corresponding point of a given point m_r should lie: in the epipolar line l_{m_r}. This condition is called the epipolar constraint. The lines $(\overline{C_l\ C_r})$, $(\overline{C_l\ m_l})$, and $(\overline{C_r\ m_r})$ are contained by the plane Π. This co-planarity can be used for solving correspondences in image matching as follows:

$$m_l{}^T E\ m_r = 0, \tag{2.8}$$

where the matrix E is a 3x3 matrix defined by a rotation a and translation between the two cameras, under the assumption that the intrinsic camera parameters are known (Hartley & Zisserman, 2004). The matrix E is called the essential matrix. It is the mapping between points and epipolar lines. Consequently, for a given point m_l in I_l its corresponding epipolar line in I_r is:

$$l_{m_r} = E^T m_l \tag{2.9}$$

Nevertheless, in practice, the intrinsic camera parameters are unknown and they have to be estimated. In this case, the mapping between points and epipolar lines can be obtained from corresponding points, with no information at all on the camera parameters, based on the fundamental matrix F. The fundamental matrix is defined in terms of pixel coordinates, whilst the essential matrix is defined in terms of camera coordinates. The fundamental matrix encodes both the intrinsic and the extrinsic camera parameters. It relates points, epipolar lines, and corresponding points as follows:

$$l_{m_r} = F^T m_l, \tag{2.10}$$

$$m_l{}^T F\, m_r = 0. \tag{2.11}$$

Further information about the estimation of the fundamental matix can be found in (Longuet-Higgins,1981; Faugeras, 1992; 1993; Zhang et al.,1995, Luong & Faugeras, 1996; Xu & Zhang, 1996; Torr & Murray, 1997; Hartley,1997; Zhang,1998; Hartley & Zisserman, 2004), among others.

A canonical stereo camera system or stereo model is illustrated in Figure 2-7. In this model, the optical axes are parallel and perpendicular to the baseline and also the x-axis coincides with the baseline. In addition, epipolar lines coincide with the x -axis. A canonical model can be reached from different camera configurations by a transformation process termed rectification. Further information about the rectification process can be found in (Loop & Zhang, 1999; Isgro & Trucco, 1999; Fusiello et al., 2000; Kang et al., 2008), among others.

The epipolar constraint is of geometric nature. Its use decreases from 2D to 1D the search process of corresponding points. If there exists uncertainty about the imposition of the epipolar constraint, the search process may consider some additional rows above and below the estimated epipolar line associated of each point.

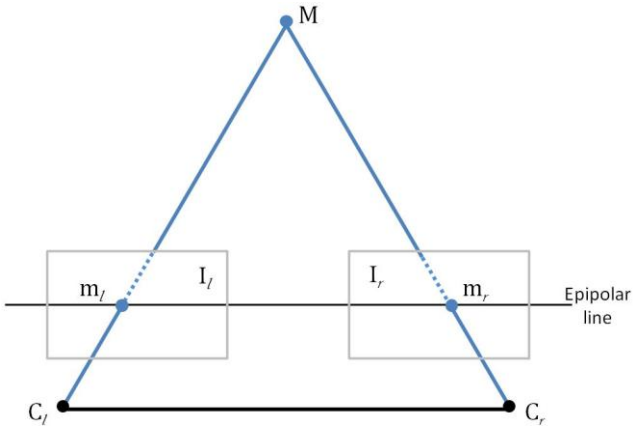


Figure 2-7 Epipolar constraint in a canonical stereo camera model.

2.2.3.2 Uniqueness Constraint

The uniqueness constraint restricts the number of possible matches to at most one (Marr, 1982). The implications of the uniqueness constraint are twofold. Firstly, each physical point is only allowed to occupy one and only one position at a given time. Secondly, it restricts the stereo image content to involve exclusively opaque objects. This constraint excludes translucent and/or transparent surfaces, since through these surfaces points of different depth are projected onto an image plane common point.

2.2.3.3 Continuity Constraint

The continuity constraint involves the cohesiveness of the matter in the physical world, and states that 3D scenes are composed by objects whose surfaces are generally smooth. It implies that adjacent points in 3D space will have adjacent projections. Consequently, depth is varying smoothly almost everywhere. Clearly, the continuity constraint does not hold near object boundaries.

The continuity constraint has also been formulated in different ways such as the Disparity Gradient Limit constraint, and the figural continuity constraint (Goulermas, 2000).

2.2.3.4 Compatibility Constraint

The compatibility constraint states that there are invariances between the stereo views of a 3D scene. Consequently, corresponding points may have similar properties.

For instance, it is reasonable to assume that corresponding points may have similar intensity values, due to a photometric invariance. In addition, it is also reasonable to assume a preservation of geometric shapes or structures due to geometric invariances. These invariances have a higher probability of appearing when the baseline of the stereo camera is short.

2.2.3.5 Scene *a priori* Information

Based on scene *a priori* information it is possible to restrict the search of corresponding points. This information may be about the disparity range of the scene, or about the shape of the object surfaces, among others. The information about the disparity range of the scene makes possible to limit the search for matching points to just a segment of the epipolar lines. On the other hand, if a model of the surfaces' shapes is available, the disparity function can be adjusted to fit such model. For instance, if the scene is composed by only frontoparallel objects, the disparity function may be restrained to have some step discontinuities and constant values elsewhere.

2.3 3D Reconstruction

The 3D reconstruction process can be addressed by a set of corresponding points. In fact, there are three cases which determine the type of the possible 3D reconstruction (Hartley & Zisserman, 2004). They are related to the amount of information available about stereo camera system and knowledge on matching points. Each one of these cases may recover sufficient information to a user, according to the particular application domain of interest. The three reconstruction cases are briefly explained below.

2.3.1 Projective Reconstruction

A projective reconstruction of a scene can be computed from two views based on image correspondences alone, without knowing anything about the calibration or the pose of the two cameras involved. The projective reconstruction is the first step to achieve the subsequent reconstructions in a stratified reconstruction approach.

2.3.2 Affine Reconstruction

The essence of the affine reconstruction is to locate the plane at infinity. This knowledge is equivalent to an affine reconstruction, and can be achieved by different means (i.e. available information), where different means may have different advantages in practice. Some of the circumstances that can be used for achieve an affine reconstruction are:

- the cameras are known to undergo a pure translational motion without change in the internal parameters;
- scene constraints, such as the identification of three point lying on the plane at infinity;
- identify in the scene the presence of, at least, three set of parallel lines with different direction, allowing the determination of the plane at infinity in the projective reconstruction;
- knowing world distance ratio of a line in the image;

  among others.

2.3.3 Metric Reconstruction

The key of the metric reconstruction is to identify the absolute conic: a planar conic lying in the plane at infinity (Hartley & Zisserman, 2004). In practice, a way to accomplish this goal is to consider the image of the absolute conic in one of the images. The image of the absolute conic is a conic in the image, which back-projection is a cone meeting the plane at infinity in a single conic, defining the absolute conic. Three sources of constraints, as well as combinations of them, are used to identifying the image of the absolute conic:

- constraints arising from scene orthogonality,
- constraints arising from known internal parameters,
- constraints arising from the same cameras in all images.

On the other hand, a jump from the projective reconstruction to a metric reconstruction is possible if ground control points are available. This is, points with known 3D locations in Euclidean world frame are given. This may be the case, for instance, when a calibration pattern is used (Zhang, 2000; 2004).

2.4 Distance Functions

In the context of the thesis, a distance function can be understood as a function of the dissimilarity, or the similarity, between two images or signals in general. Similarity and dissimilarity functions are used to produce a score of the comparison between signals, allowing a quantitative assessment between signals.

Without a loss of generalisation, a distance or dissimilarity function d can be defined as follows:

Definition 2.5: Let d be a function of the form $d: (\mathbb{N} \times \mathbb{N}) \rightarrow \mathbb{R}$. Let a , b and c be discrete values: $a, b, c \in \mathbb{N}$. A function d is considered as a metric if it fulfils the following properties:

- Nonnegativity: $d(a, b) \geq 0$.
- Identity: $d(a, b) = 0$, only iff $a = b$.
- Symmetry: $d(a, b) = d(b, a)$.
- Triangular inequality: $d(a, c) \leq d(a, b) + d(b, c)$.

In a full reference schema, such as the illustrated in Figure 2-8, a distorted image is compared against an undistorted image (i.e. a perfect quality image) producing a single score value (Wang et al., 2004).

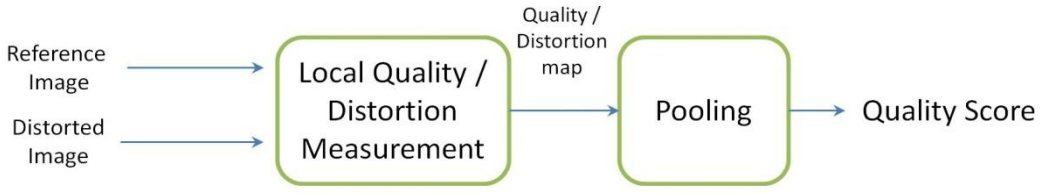


Figure 2-8 Image comparison under the full reference approach.

In the case of the stereo vision process, distance functions are used during the search of corresponding points by comparing points of the reference image, against points of the target image, within a supporting region or matching window. They are also used during an evaluation process for comparing estimated disparity maps against ground-truth maps, or rendered views based on estimated maps against original views, in order to obtain an image quality score. Without loss of generalisation, in the following notation the image planes I_l and I_r will be used for denoting the images being compared in the presented functions, and let (I_{l_i}, I_{r_i}), $(i = 1,2,\dots,n_w)$ be bivariate intensity values in the supporting windows in I_l and I_r, respectively. The image plane acts as the reference image and the image plane acts as the target image (i.e. if the goal is to establish matching points) or the distorted image (i.e. if the goal is to measure image quality).

2.4.1 Dissimilarity Functions

Most of dissimilarity functions are unbounded measures. Thus, it is not possible to have an interpretation of the maximum value. Besides, the minimum value is only interpretable when it is associated to the value zero (i.e. the identity property of a metric).

2.4.1.1 Minkowski Distance

The Minkowski distance is a general form of a dissimilarity function. It is based on pointwise signal differences. It is formulated as follows.

$$Minkowski = \left(\sum_{i=1}^{n_w} \left| I_{l_i} - I_{r_i} \right|^p \right)^{\frac{1}{p}}. \tag{2.12}$$

The Minkowski distance, also called the L^p Norm, is equivalent to the Euclidean distance, when p is equal to 2, and equivalent to the Manhattan distance when p is equal to 1. These distance functions are commonly applied to spatially located data, such as points in a Cartesian coordinate system, or places around city blocks.

2.4.1.2 Absolute Differences

The Sum of Absolute Differences or SAD, is defined as follows.

$$SAD = \sum_{i=1}^{n_w} \left| I_{l_i} - I_{r_i} \right|. \tag{2.13}$$

The SAD is one of the simplest dissimilarity functions. It is proportional to the Mean of Absolute Differences or MAD, which is defined as follows.

$$MAD = \frac{1}{\mathrm{N}} \sum_{i=1}^{n_w} \left| I_{l_i} - I_{r_i} \right|. \tag{2.14}$$

The Truncated Sum of Absolute Differences, or $TSAD$ is a variation of the SAD, on which the maximum value between the absolute difference and a threshold value ω is assumed as the difference. It is defined as follows.

$$TSAD = \sum_{i=1}^{n_w} \max\left(\left| I_{l_i} - I_{r_i} \right|, \omega \right). \tag{2.15}$$

The Mean Relative Error or MRE assumes that one observation, $-I_{l_i}-$ is a reliable value, whilst the other $-I_{l_r}-$ is not. It is defined as follows.

$$MRE = \frac{1}{\mathrm{N}} \sum_{i=1}^{n_w} \frac{\left| I_{l_i} - I_{r_i} \right|}{I_{l_i}}. \tag{2.16}$$

2.4.1.3 Squared Differences

The Sum of Squared Differences or SSD, is defined as follows.

$$SSD = \sum_{i=1}^{n_w} (I_{l_i} - I_{r_i})^2. \tag{2.17}$$

The mean of the SSD is widely known as the Mean of Squared Error, or MSE.

$$MSE = \frac{1}{\mathrm{N}} \sum_{i=1}^{n_w} (I_{l_i} - I_{r_i})^2. \tag{2.18}$$

Historically, the MSE has been widely used for optimising and assessing a variety of signal processing applications. It is the base for computing the Peak Signal to Noise Ratio or $PSNR$. The

$$PSNR = 10 \log_{10} \frac{2^{bits} - 1}{MSE}, \tag{2.19}$$

where $bits$ is the number of bits used for storing and representing the signal. The answer of the is expressed in decibels (dB), and typical quoted scores are in the range +25dB to +35dB. The $PSNR$ is undefined when the signals are identical.

Although, the MSE and the $PSNR$ fulfil many desirable mathematical properties, and have been used for assessing image quality, they do not reflect well the human perception of image fidelity and quality (Brunet et al., 2012). In practice, several signals may have different types of distortions and these functions may report the same scores when they are compared using the MSE and the $PSNR$ (Wang et al., 2002; Chandler & Hemami, 2007; Wang & Bovik, 2009; Brunet et al., 2012;). This fact is illustrated in Figure 2-9 using the Lena image (Wang et al., 2002).

Figure 2-9 Different Image distortion showing the same MSE score (Wang et al., 2002).

The original image it is shown in Figure 2-9 (a). It is compared against distorted versions of it using the MSE. Obtained scores for MSE were of 225 (apart from the 2-9 (f) on which the score was 215), disregarding the different quality than can be perceived from them by a human observer.

2.4.2 Correlation and Similarity Functions

A correlation coefficient is a measure of the linear association between two variables. It can be interpreted as a measure of independence only under the assumption of a normal distribution. In general terms, the meaning of the term correlation is twofold: connection and concordance. A measure of connection is a measure of the lack of independence between variables. With regard to the concordance, two observations of a pair of samples are concordant if both values of one pair are greater than the corresponding values of the other pair, and are discordant if for one pair one values is greater and the other smaller than for the other pair. In a correlation coefficient, a positive value indicates that the variables are concordant and a negative value indicates that they are discordant. On the other hand, the magnitude of the correlation coefficient indicates the degree of linear association.

2.4.2.1 Normalised Cross Correlation

The Pearson correlation coefficient is known as the Normalised Correlation Coefficient or NCC. The NCC takes values in the interval [-1, 1]. A value close to 1 means strong linear association, whilst a value close to zero means lack of linear association between the samples in the supporting regions. The NCC is formulated as follows.

$$NCC = \frac{\sum_{i=1}^{n_w}\left(I_{l_i} - \overline{I_l}\right)\left(I_{r_i} - \overline{I_r}\right)}{\sqrt{\sum_{i=1}^{n_w}\left(I_{l_i} - \overline{I_l}\right)^2}\sqrt{\sum_{i=1}^{n_w}\left(I_{r_i} - \overline{I_r}\right)^2}} \quad , \tag{2.20}$$

where $\overline{I_l}$ and $\overline{I_r}$ are the average of the support regions centred at the points I_{l_i}, and I_{r_i} on signals I_l and I_r,respectively. The NCC is robust against luminance or radiometric distortions between image signals.

2.4.2.2 Structural Similarity

Signals related to real imagery are highly structured, and their samples exhibit strong dependencies, especially when they are spatially proximate (Brunet et al., 2012). These dependencies carry information about the content of a 3D scene. The Structural Similarity Measure, o $SSIM$ r is based on the hypothesis that the human visual system is

highly adapted for extracting structural information from the viewing field. Thus, a high-quality image is one whose structure most closely matches the structure of the reference image. The $SSIM$ was proposed in (Wang et al., 2004), and is an improved version of the universal image quality index proposed in (Wang & Bovik, 2002). It assumes that a measure of structural information change may provide a proper approximation to perceived signal distortion, independently of the changes in luminance and contrast. Following the approach illustrated in Figure 2-8, the $SSIM$ is computed locally, for a supporting region surrounding the point of interest and a single value is collected by computing the mean of each score.

The three components of the $SSIM$ are defined as follows.

$$l\,(I_l\,, I_r) = \frac{2\,\overline{I_l}\,\overline{I_r} + k_1}{\overline{I_l}^{\,2} + \overline{I_r}^{\,2} + k_1}\ , \tag{2.21}$$

$$c\,(I_l\,, I_r) = \frac{2\sigma I_l\,\sigma I_r + k_2}{\sigma_{I_l}^2 + \sigma_{I_r}^{\,2} + k_2}\ , \tag{2.22}$$

$$s\,(I_l\,, I_r) = \frac{\sigma I_l\,I_r + k_3}{\sigma I_l\,\sigma I_r + k_3}\ , \tag{2.23}$$

where $\sigma_{I_l}^2$ is the variance of I_l, $\sigma_{I_r}^2$ is the variance of I_r, $\sigma I_l\,I_r$ is the covariance of I_l and I_r, and k_1, k_2 and k_3 are small constants ($\ll 1$). When the importance of each component is equally weighted, the resulting index is given as follows.

$$SSIM\,(I_l\,, I_r) = \frac{(2\,\overline{I_l}\,\overline{I_r} + k_1)\,(2\sigma_{I_l I_r} + k_2)}{\left(\overline{I_l}^{\,2} + \overline{I_r}^{\,2} + k_1\right)\left(\sigma_{I_l}^2 + \sigma_{I_r}^{\,2} + k_2\right)}\ . \tag{2.22}$$

The $SSIM$ fulfils the following conditions.

- Boundness: $SSIM(I_l\,, I_r) \leq 1$.
- Symmetry: $SSIM(I_l\,, I_r) = SSIM(I_r\,, I_l)$.
- Unique maximum: $SSIM(I_l\,, I_r) = 1$.

It has been extended to handle multiple scales (Wang et al., 2003) and information content weighting for perceptual image quality assessment (Wang & Li, 2011).

2.4.3 Non-parametric Distance Functions

Non-parametric distance functions are based on the local order of the signal samples rather than the intensity or sample magnitude themselves. They are robust against radiometric distortions between image signals (Zabih & Woodfill, 1994).

2.4.3.1 Rank Transform

The rank transform for a supporting region surrounding a point of interest is defined as the number of samples in the region for which the sample magnitude is less than that of the point of interest. The rank transform applied to a supporting region I_l centred at I_{l_i} is formulated as follows:

$$d(p\,,q) = \begin{cases} 1, if\ p > q \\ 0, if\ p \leq q \end{cases} \tag{2.23}$$

$$Rank\ (I_{l_i}) = \sum_{j=1}^{n_w} d(I_{l_i}\,, I_{l_j}). \tag{2.24}$$

After the rank transform is applied to both signals, the signals are compared by the SAD distance. Although the rank transform brings robustness against radiometric distortions, it also may imply some loss of information due to the compression of information: the relative ordering of samples within the supporting region is transformed into a single value (Brown et al, 2003).

2.4.3.2 Census Distance

The census transform is similar to the rank transform, but it aims to preserve the spatial distribution of the samples within the supporting region into an encoded bit string (Zabih & Woodfill, 1994). The bit string is constructed according to the comparison between the samples within the supporting region against the point of interest. A 1 bit is assigned when the magnitude of the point of interest is greater than that of the sample, and 0 bit is assigned otherwise. After the construction of bit strings, the signals are compared by the Hamming distance. The Hamming distance measures the number of substitutions between the two bit strings. It is a dissimilarity measure. Let $\widetilde{I_l}$ and $\widetilde{I_r}$ be the of bit strings associated to the supporting regions I_l and I_r , respectively. Let n be

the length of the bit strings. Then, the Hamming distance of a pair $\widetilde{I_l}$ and $\widetilde{I_r}$ is formulated as follows.

$$Hamming\ (\widetilde{I_l}, \widetilde{I_r}) = \sum_{j=1}^{n} \begin{cases} 1, if\ I_{l_j} \neq\ I_{r_j} \\ 0, if\ I_{l_j} =\ I_{r_j} \end{cases}. \tag{2.25}$$

The Census distance alleviates the loss of information associated to the Rank transform, but it implies a higher computational cost (Brown et al., 2003).

2.5 Multi-objective Optimisation

It is common to find engineering problems involving multiple objectives (Horn et al, 1994; Goldberg, 1998; Knowles & Corne, 1999; Veldhuizen & Lamont 2000; Deb et al., 2002; Veldhuizen 2003). In general, multi objective problems require two tasks: search and decision making. The search process takes place in the decision variable space, whilst the decision process takes place in the objective function space. The decision variable space, denoted as Ω, is a set of vectors in $\mathbb{R}^{\mathrm{n}}$, $\Omega = \{\vec{\mathrm{V}} \in \mathbb{R}^{\mathrm{n}}\}$. The objective function space, denoted as Λ, is a set of vectors in $\mathbb{R}^{\mathrm{k}}$, $\Lambda = \{\vec{\mathrm{U}} \in \mathbb{R}^{\mathrm{k}}\}$. In fact, the search space associated to the problem can be too large to be enumerated, or too complex, or both, in order to be explored by a conventional optimisation technique (i.e. such as linear programming or local gradient search). Moreover, these type of problems are characterised by the presence of several conflicting and incommensurables objectives which may involve difficult trade-offs during a decision making process. Thus, the presence of a single optimal solution is infrequent. Instead, there is a set of alternative trade-offs solutions. These solutions are termed as Pareto-optimal solutions. They are optimal in the sense that no other solutions in the search space are superior to them considering all involved objectives.

Without loss of generalisation a description of a Multi Objective optimisation Problem or MOP is provided as follows.

Definition 2.6:, A MOP consists in finding the vectors of decision variables $\vec{v} = (v_1, v_2, \dots, v_n)^T$ that optimises the following equation.

$$Min_{\vec{v}}\ \ g(\vec{v}) = (g_1(\vec{v}), g_2(\vec{v}), \dots, g_k(\vec{v})), \tag{2.28}$$

subject to:

$$\vec{v} \in \Omega, \tag{2.29}$$

$$h_p(\vec{v}) \leq 0 \quad o = 1, \dots, P \ , \tag{2.30}$$

$$j_q(\vec{v}) = 0 \quad q = 1, \dots, Q \ , \tag{2.31}$$

where $g_k: \mathbb{R}^n \to \mathbb{R}\ (k = 1, \dots, K)$ are the objective functions, and h_p and $j_q: \mathbb{R}^n \to \mathbb{R}\ (o = 1, \dots, P\ ;\ q = 1, \dots, Q)$ are the constraints of the problem.

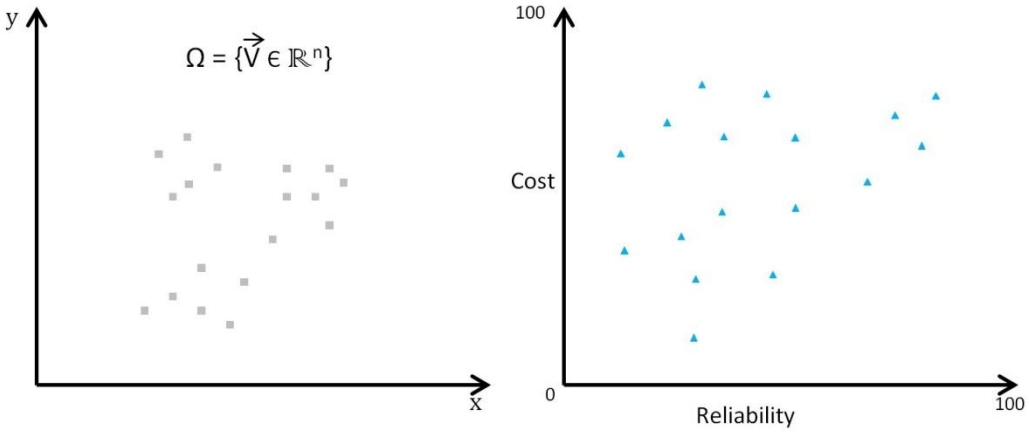


Figure 2-10 A MOP evaluation function mapping between the decision and the objective space.

The MOP evaluation function g is illustrated in Figure 2.10, for the case $n = 2$, $k = 2$, and $m = 0$. This function computes a mapping form vectors in the decision variable space, to vectors in the objective function space.

In addition, definitions of the Pareto Dominance relation, non-dominated solutions, Pareto optimal solution, Pareto Optimal set, and the Pareto front, are presented below for the sake of completeness, since they support the decision making process.

Definition 2.7: Given two solutions $\vec{v},\ \vec{u} \in\ \mathbb{R}^n$, $\vec{v}$ dominates u, denoted as $\vec{v} \prec u$, if and only if: $f_a(\vec{v}) \leq f_a(\vec{u})\ \forall\ a \in \{1, .., K\}$ and $\exists\ b\ \in \{1, .., K\}$ where $f_b(\vec{v}) < f_b(\vec{u})$.

Definition 2.8: A solution $\vec{v}\ \in\ \Omega$ is non-dominated if and only there not exist another solution $\vec{u}\ \in\ \Omega$, such that $\vec{u} \prec \vec{v}$.

Definition 2.9: A solution $\vec{v} \in \Omega \subseteq \mathbb{R}^n$, where Ω is the decision space, is Pareto optimal if it is non-dominated with respect to Ω.

Definition 2.10: Let $\mathcal{P}^*$ be the Pareto optimal set defined as $\mathcal{P}^* = \{\vec{v} \in \Omega, v \text{is Pareto optimal.}$

Definition 2.11: Let $\mathcal{PF}^*$ be the Pareto front, defined as $\mathcal{PF}^* = \{ g(\vec{v}) \in \Lambda, v \in \mathcal{P}^* \}$.

In practice, search and multiple criterion decision making are not totally independent tasks, since making some multiple criterion choices before, or during the search may alter or biases the search result. Three strategies can be identified for the approaches addressing the multiple objective problems, with regard to the relation between the search and the decision making process:

- Make decisions before to the search process.
- Search first, and then conduct decision making.
- Integrate, iteratively, the search process with the decision making.

An illustration of the Pareto dominance is presented in Figure 2-11 using $= 2$. The example illustrates a problem of minimising product costs and maximising product reliability within a manufacturing process, on which a population of solutions is plotted using evaluated criterion as a coordinate.

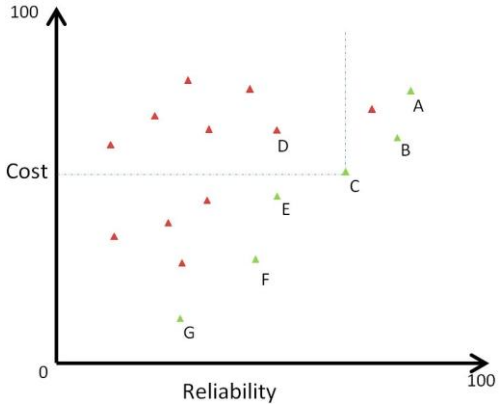


Figure 2-11 Objective function space in a two criteria problem.

It can be observed that, the solution C dominates the solution D, but it does not dominate the solution B. In fact, C and B are incomparable among them, and make part of the Pareto optimal set of the problem.

2.6 Chapter Summary

Some topics presented in this chapter, and relevant to the scope of the thesis, are summarised below.

- The stereo correspondence problem is an ill-posed problem due to the lack of information about depth and data instability. As a consequence of data instability, a small perturbation in the matching of conjugated points may produce a large error in the 3D information recovery process.
- There are three types of errors during the searching for corresponding points: mismatches, false negatives and false positives.
- The multi objective optimisation theory gives support to deal with problems involving many conflictive and incommensurable objectives. It also bring support to the decision making process when engineering problems are addressed.

CHAPTER 3.

LITERATURE REVIEW

The chapter contains a review of some approaches regarding the stereo correspondence problem, in two different stages of a 3D reconstruction process: searching an initial set of corresponding points in views generated by an non-calibrated system (i.e. prior to the imposition of the epipolar constraint), and searching for corresponding points in the entire image (i.e. after the imposition of the epipolar constraint by the rectification of the stereo image pair). In addition, once matching points have been estimated, what alternatives are for evaluating the accuracy of disparity maps. Different proposals found in the literature to deal with the above problems, among others closely related issues, are reviewed in this chapter. The content of the chapter is as follows.

Chapter contents

3.1 Classification of Stereo Correspondence Methods

The searching for corresponding points in images capturing a single scene is still attracting attention in the computer vision community (Faugeras, 1993; Xu & Zhang, 1996; Lowe, 2004; Hartley & Zisserman, 2004; Moreels & Perona, 2005, Bay et al., 2008). Historically, algorithms for stereo matching have been classified in two main

categories: template matching and feature matching (Barnard & Fischler, 1982; Dhond & Aggarwal, 1989; Goulermas et al, 2005). On the one hand, template based matching methods attempt to correlate the grey (or even colour) levels of image regions in the views under analysis, assuming that they present image irradiance similarities. Thus, most template based methods may be sensitive to illumination and contrast distortions. Their advantage consists in the generation of dense disparity measurements, since disparity can be estimated at most pixel points (Goulermas, 2000). On the other hand, feature matching based methods aims to first extract salient primitives from the images and matching them assuming a geometric invariance. These primitives are, in brief, local, meaningful, detectable parts of the image, such as interest points, corners, edges segments or contours, among others (Trucco & Verri, 1998, Schmid et al., 2000). These methods are fast since only a small subset of the image pixels are used, but they may fail if the chosen primitive cannot be reliably detected in images. Moreover, a post-processing stage is required in order to fill the gaps if only these methods are used. A more recent classification of stereo methods is based on the type of optimisation strategy used for computing matching points (Scharstein & Szeliski, 2002). In this way, a broad distinction between local optimisation and global optimisation strategies is proposed. Moreover, a taxonomy based on a set of algorithmic building blocks of stereo correspondence methods is proposed in (Scharstein & Szeliski, 2002). The proposed classification of stereo algorithms assumes a canonical stereo camera system.

An alternative classification of stereo correspondence algorithms may be based in the constraints used (Goulermas, 2000). In this chapter, stereo correspondence methods are broadly divided into methods for finding an initial set of corresponding points prior to the estimation of the fundamental matrix, and methods for finding corresponding points on the entire image after the imposition of the epipolar constraint.

3.2 Stereo Correspondence Methods for a Search in 2D

In general terms, the methods more suited to find an initial set of corresponding points are feature based methods using a pixel-based primitive (Harris & Stephens, 1988; Tissainayagam & Suter, 2004; Kerr et al., 2008). Although a more sophisticated

token may provide robustness to the initial matching process, a pixel-based feature provides the minimal enough information required, without an additional processing due to the information refinement that might be necessary in a more complex token (Park & Han, 1998).

Most of the methods for establishing initial correspondences have three main phases or steps: feature detection, feature description and feature matching (Mikolajczyk & Schmid, 2005; Bay et al., 2008). In the detection phase, a feature of interest is found and located. In this regard, there are three principles of optimality that a feature detector operator should satisfy (Canny, 1986):

- Good Detection: the detector should exhibit the highest rate of true-positive responses and the lowest rate of false-positive responses, as possible.
- Good Localisation: the reported position of the detected feature should be as close as possible to the centre of the feature in the image.
- Single Response: an image feature should generate just one response by the operator.

In addition, there are some criteria that make a particular token more useful or suited for a particular task, than others. Compactness, repeatability and distinctiveness are among these criteria (Ayache, 1991; Schmid et al., 2000; Lowe, 2004).

- Compactness implies that a token should be as concise as possible in order to avoid introducing additional complexity into the whole process.
- Repeatability signifies that a feature should be detected no matter the possible changes on imaging conditions (i.e. such as changes in scaling, rotation, illumination and 3D camera viewpoint). Repeatability may be also termed as stability in the literature (Goulermas, 2000; Olague & Trujillo, 2012).
- Distinctiveness signifies that a feature can be described by adequately complex properties in order to avoid false-negative responses by an operator.

One of the most used features for establishing initial correspondences is the corner feature. The corner related token properly fulfils the above mentioned criteria. Corners are not affected by illumination changes and are rotational invariant. The

detection of corners is useful and achievable in the context of stereo matching and fundamental matrix estimation, with or without the use of a calibration pattern (Deriche & Giraudon, 1993; Zhang et al.,1995; Xu & Zhang, 1996). In addition, corners are often more abundant in real images than other features such as edges and are considered as a fundamental feature (Tissainayagam & Suter, 2004; Coleman et al., 2007). In fact, while these feature detectors operators are usually called corner detectors, they are not selecting just corners, but rather any image location that has large gradients in all directions at a predetermined scale (Lowe, 2004).

With regard to a feature descriptor, it has to be distinctive and at the same time robust to noise, detection displacements and geometric and radiometric deformation (Bay et al., 2008). The most basic descriptor is a window of image intensities.

In the matching phase, descriptor vectors are matched between different images, by a distance measure. The dimension of a descriptor has an impact on this phase. A descriptor with a low dimension allows a fast interest point matching. However, they are in general less distinctive than high dimensional descriptor vectors, which are also more computationally demanding.

In the two following sections, some proposals on intensity based corner operators and interest point descriptors are briefly reviewed.

3.2.1 Corner Detectors

Corners detectors operators can be divided into two main types: contour based and intensity based (Shah & Jain, 1984; Deriche & Giraudon, 1993). Contour based methods recovers first image contours and then search for curvature maxima or inflection points along those contours (Deriche & Giraudon, 1990; Mokhtarian & Suomela, 1998; Awrangjeb et al., 2012). Intensity based detectors compute a function of *cornerness* defined as the product of gradient magnitude (i.e. a measure of *edgeness*), and the rate of change of gradient direction with gradient magnitude, where corners are detected by thresholding (Noble, 1988). Intensity based detectors are faster than contour based detectors, and are independent of other local features. This type of detectors receives more attention in this chapter.

A rotationally invariant operator called DET is proposed in (Beaudet, 1978). The DET operator is derived using a second order Taylor's expansion of the intensity surface. It can be interpreted as the Hessian determinant, which is related to the *Gaussian* curvature. The corner detection is based on the thresholding of the absolute value of the extrema of the operator. However, this operator does not allow accurate corner localisation (Deriche & Giraudon, 1993).

The Moravec operator selects points having a high variance between adjacent pixels in four directions (Moravec, 1977; 1979). This operator computes the local maxima of a directional variance measure over a 4x4 (or 8x8) window around the point being analysed. The sum of squares of differences of adjacent pixels was computed along all four directions, and the minimum sum was chosen as the value returned by the operator. This returning value makes it sensitive to noise along strong edges. The site of the local maximum of the values returned by the interest operator was chosen as a feature point (Dhond & Aggarwal, 1989). Some weaknesses of the Moravec operator, such as its anisotropic response and noisy response, are identified in (Harris & Stephens, 1988). The author evaluates their proposal qualitatively using real imagery (Moravec, 1980).

The Marr-Hildreth operator was proposed for locating edge points (Marr & Hildreth, 1980). The operator convolves a mask approximating the Laplacian of a Gaussian function over the entire image and labels the zero-crossings of the convolution output as edges points. The edge orientation on a zero-crossing contour is given by the direction of the gradient of the convolution output, and the edge orientation is proportional to the magnitude of the gradient convolution output. This operator, as well as modifications of it, has been used for high variance point extraction (Dhond & Aggarwal, 1989).

The Kitchen & Rosenfeld (1982) corner detector uses a measure of cornerness based on the change of gradient direction along and edge contour multiplied by the local gradient magnitude. They found that the local maximum of the proposed measure isolated corners using a non-maximum suppression process applied on the gradient magnitude before its multiplication with the curvature. In fact, their cornerness measure is the explicit representation for the second directional derivative in the direction

orthogonal to the gradient (Deriche & Giraudon, 1993). Image points are declared corners if the cornerness value meets some threshold requirements.

An operator based on a facet model approach was proposed in (Zuniga & Haralick, 1983). It assumes that image intensities can be modelled by a bicubic polynomial surface. The cornerness value is computed as the rate of change in gradient angle, and a corner is detected by thresholding.

An operator based on the Gaussian curvature principle, is proposed in (Dreschler & Nagel, 1982). Nevertheless, it was shown in (Shah & Jain, 1984), that the corner operators proposed in (Zuniga & Haralick, 1983; Kitchen & Rosenfeld,1982; Dreschler & Nagel, 1982) are mathematically equivalent. Moreover, it was shown in (Deriche & Giraudon, 1993) that these operators have corner localisation problems.

The Harris operator (Harris & Stephens, 1988), which is presented as a combined edge and corner, is based on an underlying assumption that corners are associated with maxima of the local auto-correlation function. The squared first derivatives of the image are averaged over a window. The eigen values of the resulting matrix are the principal curvatures of the auto-correlation function. If these two curvatures are high, an interest point is declared. The Harris operator is sensitive to changes in image scale (Lowe, 2004). Thus it may miss significant corners, producing false negatives, and produce false positives due to noise presence. In fact, image smoothing is required since this operator relies on spatial derivatives (Wang & Brady, 1994; Tissainayagam & Suter, 2004). Moreover, this operator may work well with the “L”-junction type of corner, which corresponds to a corner of polyhedral surface in a 3D scene, but may fail in the detection of other types of features (i.e. such as the “T”-junction which arises where three polyhedral surface meet) (Noble, 1988). The Harris operator is one of the most widely used corner operators. Variations of it, aiming for instance to compute the derivatives in a more precise way, or to reduce their computational cost without impacting their properties, among others, have been developed by different authors (Trajković & Hedley, 1998; Schmid et al., 2000). The authors compare their work against the Moravec operator in a theoretical way, and evaluate their proposal qualitatively.

A formal model of corners is presented in (Deriche & Giraudon, 1993). In fact, the presented model is developed for trihedral vertexes, on which the corners are a special type of vertex. An analytical study of the behaviour of other operators such as (Beaudet, 1978; Kitchen & Rosenfeld, 1982) is conducted based on the presented corner model. The study revealed localisation problem on those operators, and the gained information was used to design the proposed operator, which is based on the properties from the Laplacian and the measure proposed in (Beaudet, 1978). During the experimental evaluation they include synthetic images (with and without noise) and a real image, which are evaluated qualitatively.

An operator based on the observation that the total curvature of the grey-level image is proportional to the second order directional derivative in the direction tangential to the edge normal and inversely proportional to the edge strength is proposed in (Wang & Brady, 1994; 1995). The proposed operator uses linear interpolation to compute second directional derivatives. Noise is reduced by local non-maximal suppression. This operator is similar to the proposed in (Kitchen & Rosenfeld, 1982), and improves corner localisation.

In the SUSAN corner detector, each pixel in the image is used as the centre of a small approximated circular mask (Smith & Brady, 1997). The greyscale values of all pixels within the circular mask are compared with that of the centre, which is termed the nucleus. All pixels with similar brightness to the nucleus are assumed to be part of the same structure in the image. They compose a structure termed the Univalue Assimilating Nucleus (USAN), based on which the presence of a corner is determined. The USAN structure is illustrated in Figure 3-1. The intensity of the nucleus is compared against the intensity of every other pixel within the mask using a function. The function, which is based on a threshold, allows the pixels to vary slightly. An USAN corresponding to a corner is one with an area of less than a half of the total mask area. Changes in the threshold value will increase, or decrease, the quantity of corners found. The localisation of the corner is achieved by finding a local minimum in the function output (Tissainayagam & Suter, 2004). Although the SUSAN operator can handle the corners generated by different types of junctions, it may have poor repeatability (Trajković & Hedley, 1998). The authors evaluate their proposal, analytically, and qualitatively on real and synthetic images.

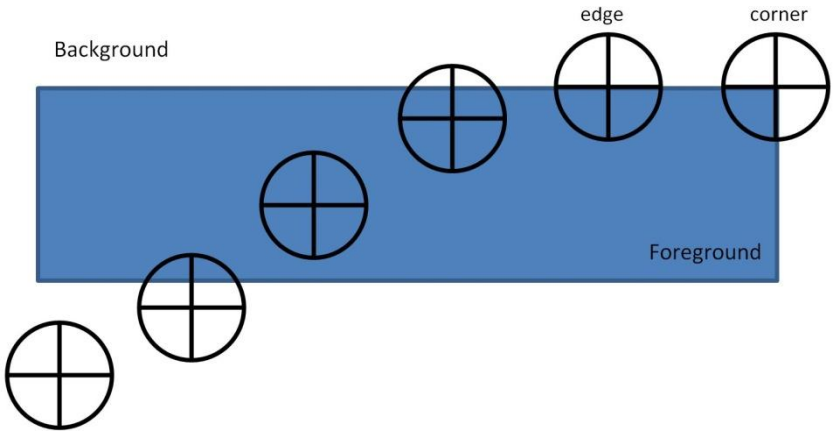


Figure 3-1 SUSAN structure for corner detection.

An operator termed Minimum Intensity Change (MIC) is proposed in (Trajković & Hedley, 1998). It is based on the variation of image intensity along arbitrary lines passing through the point of interest, within a neighbourhood. A corner is detected if the variation of image intensity along lines is high for all line orientations. The variation is found using only first derivatives. This proposal assumes that the number of corners is usually low in comparison to the image size, and uses first an intensity variation function of low computationally cost. The aim of such function is to rapidly reject no-corner points. Then, in a second step, the points identified as possible corners are processed using a higher scale and more complex intensity variation function. The corners are localised among those points that meet the thresholding requirements, in a third step based on a non-maximal suppression process. The authors compare their proposal against the SUSAN, the (Wang & Brady, 1994), the SUSAN, and the Harris operators, as well as a modified Harris operator. The localisation, the repeatability and the computational cost of operators are the evaluation criteria. The authors conclude that their proposal shows an excellent performance in terms of computational cost, and a good behaviour in terms of repeatability and localisation, whilst the modified version of the Harris operator has an excellent behaviour in terms of repeatability and good behaviour in terms of localisation (for corners related to “L”-junctions) and computational cost.

An operator to detect scale invariant interest points is presented in (Mikolajczyk & Schmid, 2001). The proposed operator uses the Harris detector at different scale-space levels, and then it selects point for which the Laplacian attains a maximum over scales. The scale at which the Laplacian is maximum is termed as the characteristic scale of the

point. This allows selecting a subset of the points detected in scale space. The operator is termed the Harris-Laplacian. The points detected by the Harris-Laplacian operator are invariant to scale, rotation, illumination changes and limited changes of viewpoint. This operator shows a better repeatability than similar operators, according to conducted evaluation on the proposal

With regard to comparisons of performance among the different corner operators reviewed in this section, there are some works available in the literature (Schmid et al., 2000; Tissainayagam & Suter, 2004; Mokhtarian & Mohanna, 2006).

The comparison presented in (Schmid et al., 2000) involves the operators proposed in (Förstner & Gülch, 1987; Harris & Stephens, 1988; Horaud et al., 1990) as well as the contour based operator proposed in (Heitger et al., 1992). The repeatability and the distinctiveness are used as evaluation criteria, where the distinctiveness was measured by entropy. They conclude that (the modified version of) the Harris operator can be considered as the one obtaining the best results.

The comparison presented in (Mokhtarian & Mohanna, 2006) includes the operators Harris, Kitchen & Rosenfeld, SUSAN as well as the original and a modified version of the contour based operator proposed in (Mokhtarian & Suomela, 1998). The repeatability (termed as consistency on the paper) and the localisation (termed as accuracy) of corners are the criteria used in the evaluation. The authors conclude that the modified version of the contour based operator outperforms the others. In regard, to the rest of operators, it can be observed from evaluation data that the Harris operator shows a better stability, whilst the SUSAN operator shows a better accuracy, according to the considered test-bed.

A comparison of operators in the context of corner tracking is presented in (Tissainayagam & Suter, 2004). The comparison involves the Kitchen & Rosenfeld, Harris, and SUSAN operators and the operator presented in (Tomasi & Kanade, 1991), which is termed KLT. The KLT operator is theoretically similar to the Harris operator and is a tracking domain specialised operator. Repeatability and localisation are used as evaluation criteria. They conclude that the Harris and the KLT operators show the best performance.

Nevertheless, there are also several factors that make difficult obtaining a final conclusion:

- There is not a canonical or standard implementation of several operators, and most of authors implement the proposals of others, without providing enough information about the algorithm parameters used. Thus, there may be some implementation details impacting on the final performance.
- Only some papers use a common image test-bed (i.e. with real images such as blocks and house, and one synthetic image), whilst others evaluate diverse images (i.e. such as paintings or drawings, which may be in practice not clear to classify between real or synthetic imagery).
- Some authors do not include a comparison against other proposals, or include in the comparison corner operators based on a different approach (i.e. contour based), and most of they evaluate their own work in a qualitative way.
- Although a qualitative evaluation may (in most of cases) properly cover aspects such as false negatives, and false positives responses by the operator, the evaluation of the corner localisation may be too expensive in terms of time and resources, and even lead to contradictory results.

3.2.2 Feature Points Descriptors

There are different options for descriptors and associated distance measures which emphasise on different image properties like pixel intensities, colour, and texture, among others. The basic idea is to detect image regions covariant to a class of transformations to compute invariant descriptors. The matching process is supported by a computed descriptor of these image regions (Mikolajczyk & Schmid, 2005).

In (Schmid & Mohr, 1997), a neighbourhood around a corner point detected by the Harris operator is described by a set of derivatives, which stable computation is achieved by a convolution with Gaussian derivatives (Koenderink & Doorn, 1987). This set of derivatives is known as the *local jet*. Differential invariants are computed from the *local jet* and stored into a vector, which is computed at different scales. Vectors are compared between them using the Mahalanobis distance. A voting algorithm and a

multidimensional indexing are used for image retrieval over a database. The conducted evaluation involves image rotation, scale changes, viewpoint variation and partial visibility. In the case of scale changes, the repeatability of the corner detector, which varies at larges scales, may affect the recognition rates.

A descriptor based on Gaussian derivatives is used in (Mikolajczyk & Schmid, 2001). The derivatives are computed at the characteristic scales of the points detected by the Harris-Lapalacian operator. Invariance to rotation is obtained by steering the derivatives in the direction of the gradient (Freeman & Adelson, 1991). A stable estimation of the gradient direction is obtained by a histogram of local gradient orientations. Invariance to the affine intensity is achieved dividing the derivatives by the steered first derivative. Descriptors are compared using the Mahalanobis distance and a dissimilarity threshold. A voting algorithm is used to select the most similar image from a database. The conducted evaluation involves image rotation, large scale changes, small viewpoint variation and partial visibility. The method is robust against large scale changes according to conducted evaluation.

The Scale Invariant Feature Transform – SIFT – descriptor transforms image data into scale-invariant coordinates relative to local features (Lowe, 1999; 2004). The resulting feature vectors *are* termed SIFT keys. A SIFT key is represented by an orientation histogram. An orientation histogram is formed from the gradient orientations of sample points within a region around the interest point. Interest points are detected by a difference of Gaussians (DOG) (Lindeberg, 1994), and edges responses are eliminated according to the curvatures computed by a Hessian matrix. The orientation histogram has 36 bins covering the 360 degree range of orientations. Each sample added is weighted according to its gradient magnitude and a Gaussian. Location with multiple peaks in the orientation histogram will have multiple keys, created at the same location but with different orientations. SIFT features are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint. In addition, the features are distinctive, which allows a single feature to be correctly matched with high probability against a large database of features, providing a basis for object and scene recognition. For image matching and recognition, SIFT features are first extracted from a set of reference images and stored in a database. A new image is matched by individually comparing each feature from the new image to this previous

database and finding candidate matching features based on Euclidean distance of their feature vectors.

3.3 Stereo Correspondence methods for a Search in 1D

After a rectification of the stereo image pair, or in the views generated by a camera system following a canonical stereo camera model, the search for corresponding points is performed in 1D, on conjugated epipolar lines. Taxonomy of stereo correspondence methods in this condition is presented in (Scharstein & Szeliski, 2002). The taxonomy is based on the constitutive modules of a stereo correspondence algorithm. These modules are:

- Matching cost computation: the matching costs for assigning different disparity hypotheses to different pixels are calculated.
- Cost aggregation: the initial matching costs are aggregated spatially over support regions.
- Disparity optimization: the best disparity hypothesis for each pixel is computed so that a local or global cost function is minimised.
- Disparity refinement: the generated disparity maps are post-processed to remove mismatches or to provide sub-pixel disparity estimates.

In addition, the taxonomy also considers a classification of stereo correspondence methods into local and global methods. Global methods deal with a disparity surface by minimising a global cost or energy function, in which the smoothness of the surface is explicit (Yoon & Kweon, 2007). Global methods are capable of producing high quality disparity estimation but with a high computational and time cost. Global methods can be classified according to the type of optimisation strategy involved. Further information about global stereo methods can be found in: (Boykov et al, 1999; Tao and & Sawhney, 2000; Goulermas et al., 2005; Min & Sohn, 2008), among others.

Some research works show that, through carefully selecting and aggregating the matching costs of neighbouring pixels, the disparity maps produced by a local approach can be more accurate than those generated by many global optimisation techniques (Wang et al., 2006a; Gong et al., 2007).

3.3.1 Local Methods

In the context of the thesis, a local stereo correspondence is one estimating the disparity assigned to each point in an independent way from the other disparities (i.e. locally, or without any smoothness term) using the Winner-Takes-All (WTA) optimisation strategy. Conventional (or non-adaptive) stereo local methods rely on the use of fixed windows in the entire image. A non-adaptive method assumes that all pixels in fixed supporting area (usually a square window centred at the interest point) are from a similar depth in the 3D scene, and therefore they have the same disparity. These methods will produce unreliable matching cost in areas near to disparity discontinuities, where the above assumption does not hold (Yoon & Kweon, 2005). Moreover, fixed window techniques, may not provide enough information in homogeneous areas, or tend to cause thin foreground objects to disappear, as well a distortion of foreground surfaces termed as fattening (Boykov et al., 1998). The inaccuracies and the distortion introduced by the use of a fixed window in the disparity estimation process are illustrated in Figure 3-2, using the Tsukuba stereo image, an estimated maps using the SAD distance with different windows sizes.

Thus, the inconvenience of fixed windows in local stereo matching is clear. On the one hand, large supporting windows tend to cover areas of different depth and are not robust against occluded regions. On the other hand, small windows are not robust against image regions with low texture (Kanade & Okutomi,1994; Tola et al., 2010). Consequently, the disparity cannot be reliably estimated based on the computed values. This is, the minimum value of the dissimilarity function between the point being matched in the reference image, and the candidates in the target image is no longer related to real disparity.

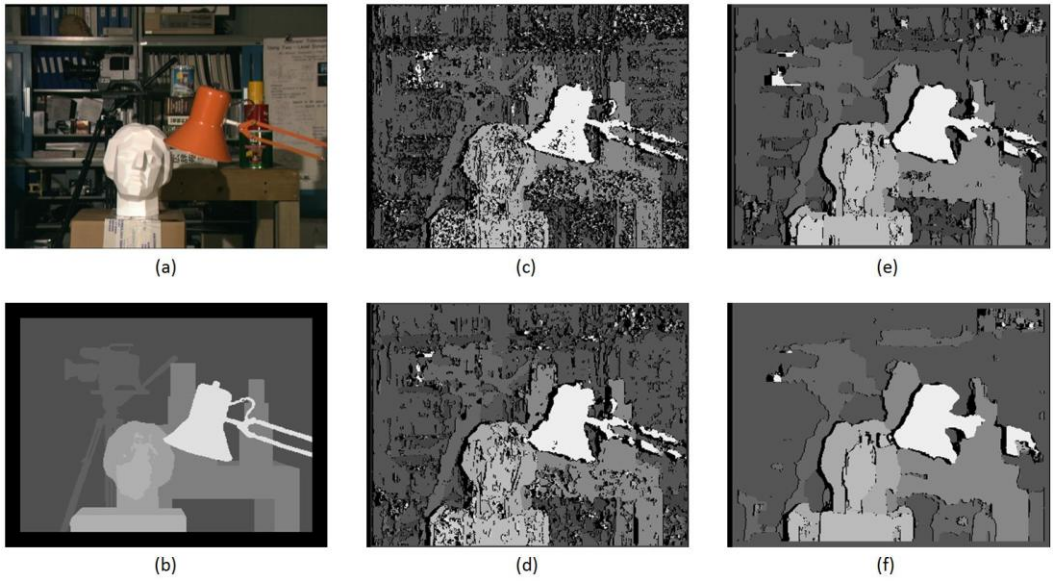


Figure 3-2 Illustration of the distortions and inaccuracies generated in conventional stereo local methods: (a) Tsukuba left view, (b) ground truth disparity map, and estimated disparity maps with windows sizes of (c) 3x3, (d) 5x5, (e) 17x17 (f) 21x21.

The above problematic give rise to the motivation for looking adaptive support regions In general terms, a support region should be large enough to include enough intensity variation for reliable matching, as well as be small enough to avoid disparity variation inside the window.

Some local methods are reviewed below according to their most distinctive characteristic.

3.3.1.1 Methods based on Multiple or Shiftable Support Regions

Symmetric multiple square windows, centred at different locations, are used in the proposal of (Fusiello & Trucco, 1997). For each pixel, nine different windows are used to aggregate the matching cost. These nine windows are illustrated in Figure 3-3 (a). The basic idea is that a window yielding a smaller matching cost is more likely to cover a constant depth area. Thus, the window with the smallest cost is retained. In this way, the disparity profile itself drives the selection of an appropriate window. Moreover, the different windows also bring some robustness again capture regions of different depth within the supporting region, as it is illustrated in Figure 3-3 (b).

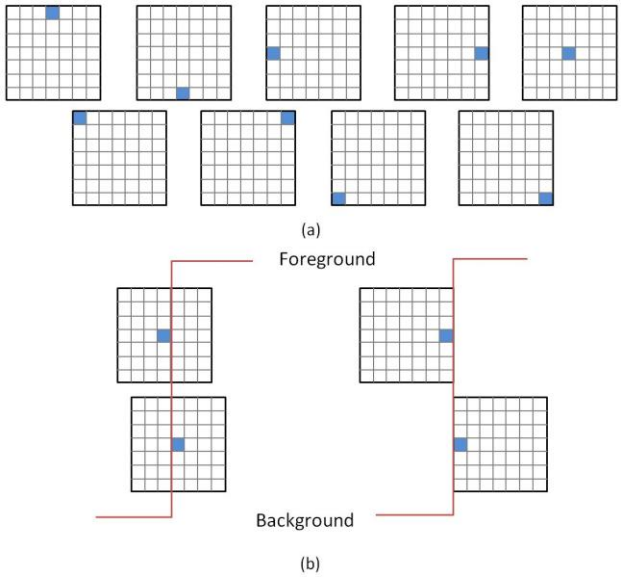


Figure 3-3 Asymmetric windows a) distribution of windows in relation to the point of interest, b) conflictive vs. convenient location of windows in relation to depth discontinuities.

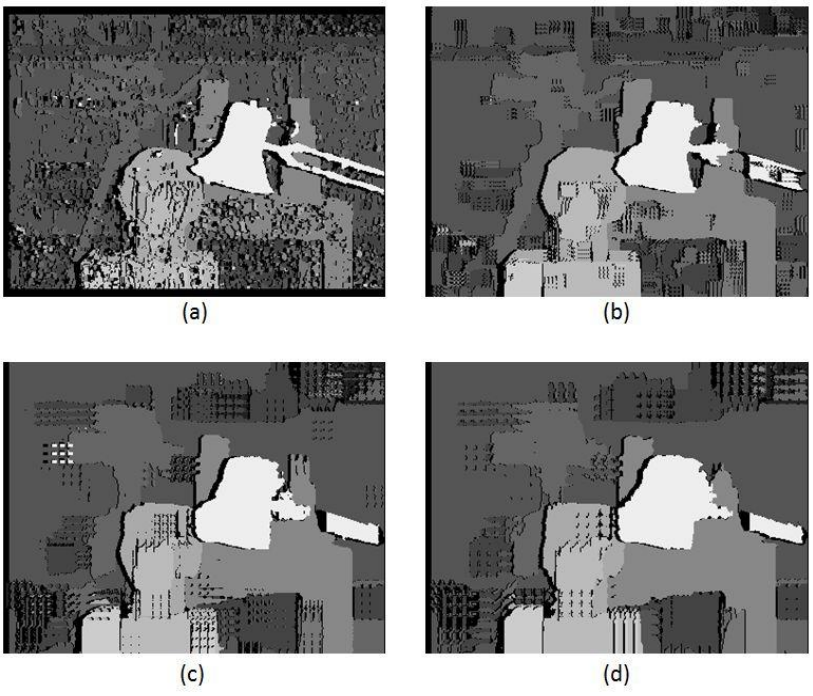


Figure 3-4 Artefacts in estimated maps using the SMW stereo method.

In practice, this proposal can be optimised in the sense that for a specific point all the multiple aggregations does not have to be computed, since each one of them has an equivalence with the aggregation computed for a window centred at a different point.

Nevertheless, this equivalency may give rise to a particular artefact on estimated disparity maps, as it is illustrated in Figure 3-4. The authors compare their approach against the adaptive windows approach of (Kanade & Okutomi,1994) using synthetic and real imagery.

The effect of shiftable-window approach can be achieved using a box filter followed by a min filter (Scharstein and Szeliski, 2002). Since a min filter is also separable, two additional rendering passes are used to compute shiftable-window results based on the square-window ones.

3.3.1.2 Methods based on Adaptive Size or Shape

A method for selecting an appropriate window for each pixel by changing the size and shape of a window adaptively is presented in (Kanade & Okutomi,1994). The adaptation is based on the local variation of intensity and initial disparity estimation. However the method is computationally expensive and highly dependent on the initial disparity estimation. Moreover, the shape of the support window is constrained to be rectangular, which may be not appropriated for pixels near to arbitrarily shaped depth discontinuities (Boykov et al., 1998; Veksler, 2002; Yoon & Kweon, 2005).

A method for choosing an arbitrarily shaped connected window is proposed in (Boykov et al., 1998). The method performs a complex plausibility hypothesis based on testing and computes a window varying at each pixel. The proposed method depends linearly form the number of pixels and the disparity range. A final maximum size window is selected form the different hypothesis of a particular point. The proposal detects occluded points explicitly. The hypothesis model is capable of handling the variation of gain and bias between stereo images. Authors compare their proposal quantitatively against the adaptive window proposal of (Kanade & Okutomi,1994) and the methods proposed in (Hanna,1974; Okutomi & Kanade, 1993; Cox et al.,1996) using synthetic and real imagery with ground-truth, and qualitatively on real imagery without ground-truth.

An adaptive shape method capable of construct non-rectangular windows is proposed in (Veksler, 2002). The window can be adapted in an efficient way using a modification of the minimum cycle algorithm from the graphs theory. However, the

computation of the window requires too many parameters (Min & Sohn, 2008). The proposal is compared against a fixed window method, the adaptive window method of (Okutomi & Kanade , 1993), and the graph-cuts based global method of (Boykov et al, 1999). Real imagery, with and without disparity ground-truth are used for quantitatively and qualitatively evaluating the proposal.

A method based on edge detection is presented in (Wang, 2004). The method determined iteratively the window size and shape according to intensity variations, which are determined by edge detection. They tried different edge operators without obtaining significantly different results. In addition, an extension to the rank measure is proposed for computing the matching costs. The extension uses two thresholds (t_{low} and t_{high}) and five ordinal values area assigned according to the comparison of the pixels within the adapted window with the central pixel.(p) It is formulated as the function $compare'$ as follows:

$$compare'(p\,,q) = \begin{cases} Smallest, if\ (p-q) < -t_{high} \\ Smaller, if\ -t_{high} \leq (p-q) < t_{low} \\ Equal, if\ -t_{low} \leq (p-q) \leq t_{low} \\ Bigger, if\ t_{low} \leq (p-q) \leq t_{high} \\ Biggest, if\ (p-q) > t_{high} \end{cases} \tag{3.1}$$

Thus, a transform of the adaptive window for the reference image is compared against the transformed windows of the target image, assigning a point for each coincidence in the ordinal level between points. The proposed method is quantitatively compared against local and global methods using the Middlebury's evaluation methodology using the first version of the online benchmark (Scharstein & Szeliski, 2002). The weakness of the proposed method is the restriction of the adapted windows to be rectangular (Gerrits & Bekaert, 2006).

An adaptive shape method based on colour similarity and connectivity is proposed in (Zhang et al., 2009a). A locally adaptive upright cross is decided upon the colour similarity, defining an initial support skeleton for the anchor pixel. Then, an arbitrarily shaped support region is dynamically constructed in the cost aggregation step reusing the previously computed neighbouring cross configuration for points in the reference and target views. An orthogonal integral image technique is used in order to accelerate matching cost aggregation over an arbitrarily shaped 2D support region by

decomposing the conventional aggregation into two orthogonal 1-D integrations. Disparities are refined by a local-high confidence voting scheme considering the support region. In this way, the disparity is assigned based on a disparities distribution peak. This assignment can be seen as a piecewise smoothness regularisation.

3.3.1.3 Methods Based on Adaptive Weights

Adaptive weight approaches use a general support window, (i.e. a rectangular fixed window) and assign adaptive weight by some compact operations (Gu et al., 2008). In this way, adaptive weight methods adapt the influence of each pixel during the disparity estimation process.

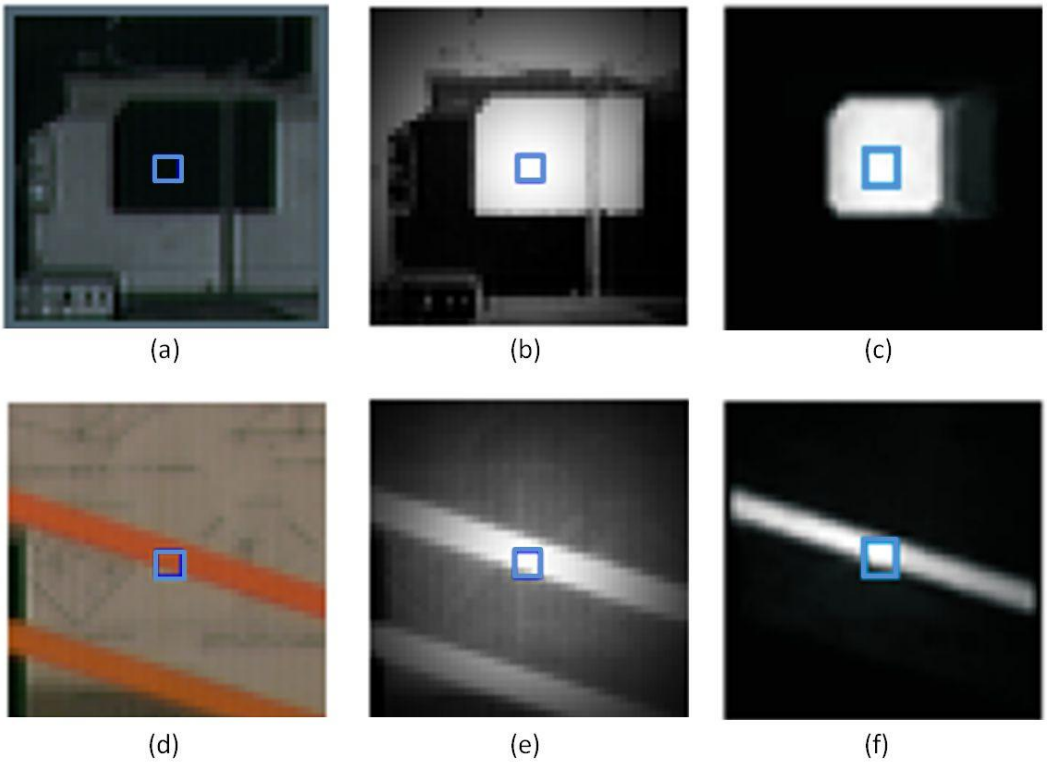


Figure 3-5 Adaptation of weights: (b) and (e) in (Yoon & Keon, 2005), (c) and (f) in (Hosni et al., 2009).

In (Yoon & Kweon, 2005) the support weight of each pixel in the window is calculated based on the gestalt grouping properties of the human visual system: similarity and proximity (Ahuja & Tuceryan, 1989). Thus, the support weight of a point within a fixed window is computed based on both the colour similarity and the Euclidean distance to the centre pixel. The similarity between two pixel colours is measured in the

CIELab colour space, which provides a three-dimensional representation for the perception of colour stimuli. The matching cost is computed considering the support-weights in both the reference and the target images. The authors compare their proposal against other local methods, using the Middlebury's evaluation methodology, and the first version of the online benchmark of it, composed by the Tsukuba, Venus, Sawtooth and Venus images (Scharstein & Szeliski, 2002). Moreover, according to experimental results, this approach can produce disparity maps comparable with those generated using global optimisation techniques (Wang et al., 2006). Nevertheless, the computation of the adaptive support weight, and the large size of the window (33x33), makes this proposal computationally expensive (Richardt et al., 2010). In addition, the weight assignment strategy may ignore the structure of the image patch, giving weights to points of similar colour but with from a different depth (Hosni et al., 2009). This is illustrated in Figure 3-5 using the Tsukuba stereo image. It can be observed in Figure 3-5 (b), that in relation to the image patch show in Figure 3-5 (a), weights are assigned to background pixels.

The colour based segmentation proposed in (Comaniciu & Meer, 1997) is used in (Gerrits & Bekaert, 2006) for segmenting the reference image Pixels within a fixed window, and belonging to the segment of the central pixel are weighted with a large value, whilst pixels outside the segment are weighted with a small value in order to reduce (but not neglecting) their influence during the matching process. This strategy is motivated as a response to the produced over segmentation in the reference image, which may cause that supporting region do not provide sufficient information for aggregation. In addition, although it is assumed that depth boundaries may coincide with colour segments, it does not implies that adjoining segments should lie in different depths. In this way, segments do not have to be explicitly modelled during the matching cost computation step. A variation of the moving average algorithm is developed and applied during the matching cost computation in order to reduce the impact of the large size widows used, on the time required by the proposal. The proposal is compared against local and global methods using the Middlebury's evaluation methodology (Scharstein & Szeliski, 2002). The weakness of the method is that the aggregation strategy is not symmetrical, since it relies only in the segmentation of the reference image (Tombari el al., 2007).

In (Tombari el al., 2007) the segmentation is applied in both the reference and the target images aiming a symmetric aggregation. In fact, it can be considered as a variation of the method proposed in (Gerrits & Bekaert, 2006) by a modification of the weights assignment function. In this function, the relevance of the pixels belonging to the segment is the same, whilst the relevance of the pixels outside the segment is varying according to a colour distance measured in the RGB colour space. The proposal is compared against the methods proposed in (Yoon & Kweon, 2005; Gerrits & Bekaert, 2006) using the Middlebury's evaluation methodology (Scharstein & Szeliski, 2002).

A method based on the idea that the distinctiveness, not the interest, is the appropriate criterion for feature selection under the ambiguous local appearances of image points is proposed in (Yoon & Kweon, 2007). In this proposal, a similarity measure based on the work of (Tomasi & Manduchi, 1999) is presented. The proposed similarity measure, termed the Distinctive Similarity Measure (DSM) is essentially based on the distinctiveness of image points and the dissimilarity between them. These properties are closely related to the local appearances of image points: the distinctiveness of an image point is related to the probability of a mismatch while the dissimilarity is related to the probability of a good match. The DSM is based on a probability model. Thus, for two image points, in the reference and the target images, respectively, the probability of obtain a correct match rises when they are distinctive, in each image, and similar between them. The method requires a threshold, which impacts on the density of estimated disparity maps. The proposal is compared against local and global methods using the Middlebury's evaluation methodology (Scharstein & Szeliski, 2002).

A method considering two steps in the matching process is proposed in (Gu et al., 2008). The steps are termed initial matching and *disparity calibration.* The step termed disparity calibration is discussed in the Section 3.4 of the chapter. The initial matching step is based on a simplified version of the adaptive support weight (Yoon & Kweon, 2005) using the RGB colour space, and the modification of the rank transform proposed in (Wang, 2004). The proposal is compared against local and global methods using the Middlebury's evaluation methodology (Scharstein & Szeliski, 2002).

A method based on colour segmentation is proposed in (Hosni et al., 2009). It is assumed that connected points, of similar colour, share the same disparity. A geodesic distance is computed from the central pixel to all pixels within a square window. The geodesic distance is low if there exists a path between the central point and the other point within the window, along which the colour varies only slightly. A path is defined as a sequence of spatially neighbouring points. Colour similarity is compared by the Euclidean distance on the RGB colour space. A higher support weight is given to pixels of low geodesic distance. Thus, these points will have a larger influence during the matching process. The connectivity principle is illustrated in Figure 3-5 (c) and 3-5 (f), regarding the image patch shown in Figure 3-5 (a) and 3-5 (d), respectively, of the Tsukuba stereo image. It can be observed that the assigned weights are restricted to neighbouring points from a same surface in 3D space. The authors consider this connectivity principle as segmentation. The proposal is quantitatively compared against other local approaches (Yoon & Kweon, 2005; 2007; Tombari el al., 2007; Gu et al., 2008), as well as to the global method proposed in (Min & Sohn, 2008), using the Middlebury's evaluation methodology (Scharstein & Szeliski, 2002). A similar idea to the geodesic distance was presented in (Darrel,1998) by a radial cumulative transform.

A multiresolution method using curvelets and adaptive support weighting is presented in (Mukherjee et al., 2010). Curvelets decompose the image into a number of scales and orientations. Matches are found in each of these scales and orientations for each point. The best match is determined by comparing correlation values and left-to-right checking. Curvelets via wrapping are combined with adaptive support weights established upon greyscale images in order to reduce the fattening effect. An initial estimation is performed in the lowest scale, and it is improved at each scale using different orientations of curvelets. Disparity values obtained in the initial disparity map restrict the subsequent range of search. The method is evaluated using the Middlebury's methodology.

3.3.1.4 Methods based on Specialised Hardware and/or Near Real-Time Performance

Local methods are suited to be implemented with a real-time performance, by exploiting the computational power offered by Graphics Processor Units (GPUs), or even

in the Central Process Units (CPUs) by sliding windows techniques (Mühlmann et al.,2002). The performances of these real-time algorithms highly depend on the cost aggregation approaches used (Wang et al., 2006a).

A method aiming a good localisation of object boundaries is (Hirschmuller, 2001; Hirschmuller et al., 2002). It is developed in the context of high-level object based tasks in a tele-operated mobile robot environment. The method, which is termed Multiple Windows Multiple Filters (MWMF), runs on standard computer hardware (Hirschmuller, 2003).

An adaptive method based on a binary window is presented in (Gupta & Cho, 2010a). The proposed method considers two main steps: initial disparity estimation and disparity refinement. The initial disparity estimation is obtained by an adaptive size and shape window. The region support is adapted based on colour similarity of points in a fixed region around the point of interest in the reference image (i.e. a 33x33 window). The similarity is measured by the Euclidean distance in the CIELab colour space. Similar points are marked, and considered for matching. Such points are termed as the active matching region. The matching cost is computed by searching the number of pixels in an active matching region having an intensity difference under a threshold value. A disparity value is selected in order to maximise such count, and it is assigned for all the points composing the active matching contour. In this way, intensity values are not used directly, reducing the impact of radiometric differences. A point may belong to more than one active matching region. If such point is associated to more than one disparity value, or if it lacks of a disparity value, the active matching region is determined taking it as the interest point (i.e. using a window of 15x15). In the disparity refinement step, neighbouring points of similar colour are forced to share the same disparity. The proposed method is compared against other real-time methods using the Middlebury's evaluation methodology.

Two windows of fixed size, one large and one small are used in the correlation based method proposed in (Gupta & Cho, 2010b). The large window (i.e. 9x9) handles low textured areas, whilst the small window (3x3) handles depth discontinuities. The proposed method consist of four main steps: an initial disparity map is computed using the two windows, matching, unreliable matches are removed by left-to-right consistency

checking, removed points in the disparity map are interpolated, and finally, the disparity map is refined in order to improve accuracy at depth boundaries using the reference image. The large size window is used first, and the disparity of the two adjacent points is taken into account in order to estimate disparity at a point in textureless areas. The small window is used at points near to depth discontinuities with a restricted disparity range. The interpolation step considers validated disparities at 8 neighbouring points, as well associated intensities in the reference image, assigning the disparity of the most similar point. The disparity refinement considers the reference image colour without performing image segmentation (Gupta & Cho, 2010b). The proposed method is compared against other real-time methods using the Middlebury's evaluation methodology.

A method with near real-time performance is presented in (Hosni et al., 2010). It is, in essence, a GPU based version of the method proposed in (Hosni et al., 2009). In this case, the geodesic distance is approximated by the Borgefor's algorithm (Borgefors, 1986). The main idea is to use the adaptive support weight windows for generating an explicit over segmentation, dividing the reference image into disjoint regions of homogeneous colour. It incorporates the sliding window technique compatible with segmentation based weight support presented in (Gerrits & Bekaert, 2006). In this way, the computational performance is no longer dependent of the size of the match window. The proposal is compared against the proposal of (Hosni et al., 2009) in terms of accuracy, by the Middlebury's evaluation methodology, as well as in execution time. A qualitative analysis is also conducted on real imagery without disparity ground-truth data.

A reformulation of the method of (Yoon & Kweon, 2005) is presented in (Rhemann et al., 2011; Hosni et al., 2012). The proposed method uses the dual-cross-bilateral filter with Gaussian weights implemented over a GPU. The Dual-Cross-Bilateral (DCB) filter is a variation of the bilateral filter, which smoothes an image with respect to edges in a different image (Paris et al., 2008; Paris & Durand, 2009). The approach explores a dichromatic based matching cost in the CIELab colour space, motivated by the memory limitations of the used platform. The dichromatic version of the proposal outperforms the accuracy of the monochromatic version, which gives poor results at disparity boundaries. Nevertheless, the dichromatic version cannot be considered as of real-time performance (Hosni et al., 2012). The proposal is compared against other real-time approaches (Gong & Yang, 2005; Wang et al., 2006b) using the Middlebury's

evaluation methodology (Scharstein & Szeliski, 2002). The proposal is also qualitatively evaluated on stereo real videos, and quantitatively on stereo synthetic videos containing Gaussian noise.

3.3.2 Global Methods

In the context of the thesis a global method is one optimising an energy function by a non-trivial optimisation strategy (i.e. dynamic programming, graph cuts, belief propagation, among others). Optimisation strategies based on, dynamic programming (Birchfield & Tomasi., 1998), and graph cuts (Boykov et al., 1999; Kolmogorov & Zabih, 2001, 2002) are widely adopted in stereo global methods. Conventional dynamic programming approaches perform optimisation in one dimension for each scanline individually, which commonly leads to streaking artefacts. The graph cuts strategy cast the searching of corresponding points into a finding the minimum cut in a graph, whilst the belief propagation strategy iteratively send messages between neighbouring nodes on the four connected image grid for minimising the global cost. Global energy functions consider at least two terms: a data term, and a smoothness term. The data term is related to the similarity of colours or intensities between points in each stereo view (i.e. regarding the compatibility constraint). The smoothness term is related to the soft changes on assigned disparities to neighbouring points (i.e. regarding the continuity constraint). Aditional terms can be used for penalising occlusions, among others.

Some global stereo methods are briefly reviewed in this Section, doing a distinction between segmentation and non-segmentation based methods, for the sake of convenience.

3.3.2.1 Global Methods based on Segmentation

A basic global matching criterion is proposed in (Tao & Sawhney, 2000). It states that if estimated disparities are correct, a rendered view based on them, should look similar to a real view. Moreover, image segmentation is exploited aiming to achieve disparity smoothness and delineated disparity boundaries. In particular, a colour based segmentation (Comaniciu & Meer, 1997) and a neighbouring disparity hypothesising algorithms are used. Disparities at each segment are modelled as a plane surface plus small variations for each point. A disparity value is hypothesised for a segment, based

on disparities of neighbouring segments, and it is tested by view rendering. The process is repeated until convergence is found or until a certain number of iteration has been executed. In this way, smoothness is enforced in homogeneous colour regions, whilst it is possible to infer reasonable disparities for unmatched areas. The proposed method was qualitatively evaluated using two outdoors stereo image pairs, and compared against a correlation based local method. Authors pointed out that the accuracy of the used warping algorithm may significantly affect the obtained disparity map.

In (Bleyer & Gelautz, 2004) colour segmentation is used in order to handle untextured regions and localisation of depth boundaries. Each segment is modelled as a plane. Segments are grouped in layers approximated by the same planar equation, and the scene is represented as a collection of layers. Segmentation is performed by a segmentation method which combines the mean shift procedure and information provided by an edge map (Christoudias et al., 2002). A sparse initial disparity map is computed by a local method exploiting reference image segmentation. The local method uses a small window (3x3) and SAD as matching cost. Estimated disparities are validated by the bi-directional constraint, and segments with a density of valid points lager than 50% are considered as reliable. The search space is reduced into those segments within the interval of maximum and minimum validated disparities. The least squared error method is used to derive a plane equation for each segment. Segments are projected into a 5-dimensional feature space. Segments of a same surface are clustered, and the members of a cluster compose a layer. Computed planes are used for layer extraction. Layers are refined by warping the reference image to the target image, and considering colour dissimilarity between warped and real view. A Z-buffering enforces visibility and allows the detection of occluded points during the warping operation. A greedy algorithm is used to optimise a cost function considering occlusion and discontinuities. The method was quantitatively evaluated using the first version of the Middlebury's online benchmark, and is also present in the second version of it (Scharstein & Szeliski, 2012). A qualitative evaluation was performed by analysing a 3D reconstruction of a self-recorded scene without disparity ground-truth data.

The method proposed in (Klaus et al., 2006) models the 3D scene as a set of planar disparity planes. It uses the mean-shift segmentation algorithm proposed in (Comaniciu & Meer, 1997; 2002). The quantity of considered planes is reduced by

extracting a set of disparity planes allowing a representation of scene's structure. A local method, with a self-adaptive dissimilarity measure combining a SAD and a gradient based measure, is applied in order achieve such reduction. Reliable disparities verified by left-to-right checking are used to derive a disparity plane. A disparity plane is assigned to each segment considering horizontal and vertical slant. Disparity planes are refined, based on aggregation of matching cost for all points inside the segment. Disparity plane labelling is optimised using loopy belief propagation where the messages are passed between adjacent segments. The proposed method is evaluated using the Middlebury's methodology.

A method modelling 3D scenes as a collection of few smooth surfaces is proposed in (Bleyer et al., 2010). In addition, it assumes that points with a similar appearance are likely (but not forced) to lie on the same 3D surfaces, as well as photo-consistency is held between stereo views. The method uses a pixel-wise Markov Random Field formulation assigning each pixel to a 3D surface, which is modelled as a B-spline. Moreover, the segmentation of the reference image, which is computed by the mean-shift algorithm (Christoudias et al., 2002), is taken as a soft constraint. In this way, the method may recover itself from initial segmentations errors. Occlusion is handled asymmetrically, and considering slanted surfaces by avoiding points from a same surface to occlude each other. The used energy function, which is composed by five terms (i.e. data, smoothness, soft segmentation, curvature and quantity of surfaces), is optimised using the fusion move approach of (Lempitsky et al., 2007). In fact, surface assignments, instead of assignments of pixels to disparity values are optimised. Surfaces assigned are initialised based on the estimation achieved by a dynamic programming based stereo method (Bleyer & Gelautz, 2008). The proposed method is evaluated in an intra and inter-technique approach using the Middlebury's methodology, in order to highlight the relevance of the different terms of the energy function.

A method for joint stereo matching and object segmentation is presented in (Bleyer et al., 2011). A higher-order Markov Random Field is used for image segmentation, which is used as a *soft* constraint (i.e. allowing colour deviation within a segment). A 3D scene is represented as a collection of visually distinct and spatially coherent objects, where each object is characterised by a colour model, a 3D plane approximating object disparity distribution and a 3D connectivity property. It is assumed

that each object is compact and connected in 3D, that all parts of an object share a similar appearance, and that a scene is composed by a few large objects. The 3D connectivity property states that disconnected 2D regions in an image may belong to the same object only if they are separated by an occluding object with a larger disparity. This property makes possible disparity estimation and assignment for disconnected background surfaces. A global energy function is optimised by a fusion move algorithm (Lempitsky et al., 2007). The method starts with an initial disparity map and an object map, and iteratively fuses the information provided by these maps. The proposed method is quantitatively evaluated using Middlebury's methodology, and qualitatively by outdoor and indoor imagery without disparity ground-truth data.

3.3.2.2 Global Methods non-based on Segmentation

A method based on constrained non-linear optimisation of a continuous disparity surface defined parametrically using radial function basis is presented in (Goulermas et al., 2005). The method is capable of explicitly incorporating a variety of types of a priori scene information by handling arbitrary constraints on the sought parameters. In this way, the optimisation of constraints and objectives is kept separated. The method uses a block decomposition scheme where the reference view is decomposed to a large number of blocks, so a large scale parallelisation becomes feasible, whilst inter-block smoothness is ensured during the optimisation cycles by enforcing block boundary continuity. The method requires an initial disparity map, which is achieved by a local method. Authors pointed out that final disparity results may be sensitive to the initial estimation. The proposed method is qualitatively evaluated, and compared against local and global stereo methods, using well known outdoors imagery (i.e. pentagon, parking meter, and shrubs).

A method reformulating the stereo correspondence problem as a large scale Linear Program presented in (Taylor & Bhusnurmath, 2008). The Linear Program can be solved using interior point methods. The presented approach is intended for situations where the displacement between frames is considerably large. The method uses an energy function considering data and smoothness terms. In addition, the smoothness term is based on disparity gradient and disparity Laplacian aiming to account planar but not front-parallel objects: the gradient based term looks for a piecewise constant model

of the disparity solution (i.e. the estimated disparity map), whist the Laplacian looks for piecewise linear model of the solution. With regard to the data term, the method computes a lower bound of the matching cost of each point by a convex hull. Matching costs are computed using the adaptive window method proposed in (Yoon & Kweon, 2005). Occluded points are explicitly detected by left-to-right checking, as well as using disparity discontinuities found in both stereo images (Sun et al., 2005). Disparity values are assigned to occluded points due to the use of the Laplacian term. The energy function is rewritten in a matrix form and is optimised using an interior point log barrier method (Boyd & VandenBerghe, 2004). The method is evaluated using the Middlebury's methodology. The computation of disparity maps for the Middlebury's benchmark took around nine minutes.

A GPU based method is presented in (Mei et al, 2011). It is based on matching cost combining absolute differences and the census transform. Absolute differences are computed on the RGB colour space, and added to census based cost after a transformation to the [0, 1] range. The combination of these two measures is motivated by the observation that the census cost is prone to produces wrong matches in regions with repetitive local structures, whilst absolute differences cannot properly deal with large textureless regions. Matching costs are aggregated over cross-shaped support regions (Zhang et al, 2009a). The support region is constructed based on colour similarity, connectivity and a maximum width. It uses a scanline optimisation framework based in a semiglobal approach with reduced path directions (Hirschmuller, 2005): two along horizontal and two vertical directions. It incorporates a multi-step disparity refinement process involving: iterative region voting, interpolation, depth-discontinuity adjustment and sub-pixel enhancement. The method is quantitatively evaluated using the Middlebury's methodology, and qualitatively evaluated using two stereo sequences. Authors point out that some artefacts arise in the estimation around depth borders and occluded regions, due to problems with the construction of the support region.

The semi global matching method proposed in (Hirschmuller, 2005) sums for each pixel the costs along 1D paths from several (i.e. 8 or 16) directions. Non-straight paths are implemented by going one step horizontally or vertically, followed by one diagonally step. Its pixel-wise cost matching is based on Mutual Information (Viola & Wells, 1995), achieving robustness against radiometric differences and reflections.

Mutual information is defined from the entropy of two images, as well as their joint entropy, which can be calculated as sum of data terms that depend on corresponding intensities. A hierarchical calculation of mutual information as matching cost is suggested, using recursively the up-scaled disparity image which has been calculated at half resolution, as the initial disparity. Authors highlight that the proposed method is of linear complexity in terms of the image size, and disparity range (if intermediate costs are properly used). An energy function of three terms is used, where the third term (besides data and smoothness) is aimed for preserving discontinuities. Disparity peaks (i.e. outliers) are detected by segmentation, where only segments over a threshold size are allowed. Occluded points and mismatches are differentiated by left-to-right checking, and arising holes are filled by discontinuity preserving interpolation. In addition, sub-pixel refinement is performed. The proposed method is quantitatively evaluated and compared using the Middlebury's methodology, and qualitatively evaluated using indoor imagery of structured environments.

3.4 Pre and Post-processing Procedures Related to Stereo Correspondence

3.4.1 Pre-processing Procedures

One of the pre-processing methods more commonly used is the segmentation of the stereo images. This pre-processing procedure is mainly conducted by global stereo methods. In this regard, the segmentation algorithm proposed in (Comaniciu & Meer, 1997), or a variation of it, is widely used.

The mean-shift colour segmentation proposed in (Comaniciu & Meer, 1997; 2002) is essentially defined as a gradient search for maxima in a density function defined over a high dimensional feature space. The feature space includes a combination of spatial coordinates and associated attributes considered during the analysis. It incorporates edge information. The proposed method is based on a nonparametric technique for estimation of the density gradient (Cheng, 1995). It can work at different segmentation resolutions (i.e. under-segmentation, over-segmentation, and quantisation), being the desired resolution specified by a user via three parameters:

the radius of the search window, the smallest number of elements required for a significant colour, and the smallest number of contiguous pixels required for a significant image region. The *L*u*v** colour space was used in order to maintain the isotropy of feature space. The analysis of the feature space is performed autonomously due to the use of image domain information. The proposed segmentation algorithm handles grey level images as colour images having only the lightness coordinate.

The information provided by an edge magnitude/confidence map is incorporated into a mean shift based colour image segmentation approach in (Christoudias et al., 2002). The aim of the method is to identify regions with weak but sharp boundaries, in order to provide a more accurate input for segmentation based high level tasks. Image segmentation and edge detection are combined since are considered as complementary in nature: image segmentation focuses on global information and labels the input of homogeneous regions, whilst edge detection focuses on local information and labels the pixels which are assumed to be located in discontinuities. It is pointed that although in principle both operations should give the same results, in practice results differ significantly since local and global evidence may lead to different conclusions. In particular, the proposed algorithm combines the segmentation approach presented in (Comaniciu & Meer, 1997) plus the gradient based edge detector with embedded confidence proposed in (Meer & Georgescu, 2001). The proposal was qualitatively evaluated.

3.4.2 Post-processing Procedures

A post-processing stage of disparity maps may involves estimation errors detection, occlusion handling, disparity map smoothing, and subpixel interpolation, among others.

With regard to the errors detection, methods rely in a confidence measure. The variance of the values produced by the dissimilarity function is used in (Fusiello & Trucco, 1997) as a measure of confidence. The relation among the three lowest values of the dissimilarity function are used in (Mühlmann et al.,2002) in order to define a uniqueness of the minimum criterion. The lowest value defines a threshold above where the third smallest value should lie. Moreover, the relation between the lowest value and the second one can be used following the equation:

$$ratio = \frac{second\_lowest - lowest\_value}{lowest\_value}, \tag{3.2}$$

where a low $ratio$ value indicates possible problems regarding the reliability of the estimated disparity. Nevertheless, the thresholds for comparing these values should be empirically set, according to the application domain (Hirschmuller, 2001; Hirschmuller et al., 2002).

Most of stereo local methods (that do not model occlusion explicitly) handle the detection of occluded points by the bidirectional or left-to-right constraint. It considers both the dissimilarity or correlation functions computed when the left and the right views are acting as the reference and the target images, respectively, and vice versa (Fua, 1993). The disparity is considered to be reliable, if the minima of both functions coincide in the same disparity. Although this process is effective in filtering out occluded pixels as well as mismatches, the bidirectional constraint is a heuristic in the sense that the coincidence of minima does not implies the estimated to be true. The bidirectional constraint is illustrated in Figure 3-6 for points belonging to the background of a synthetic scene.

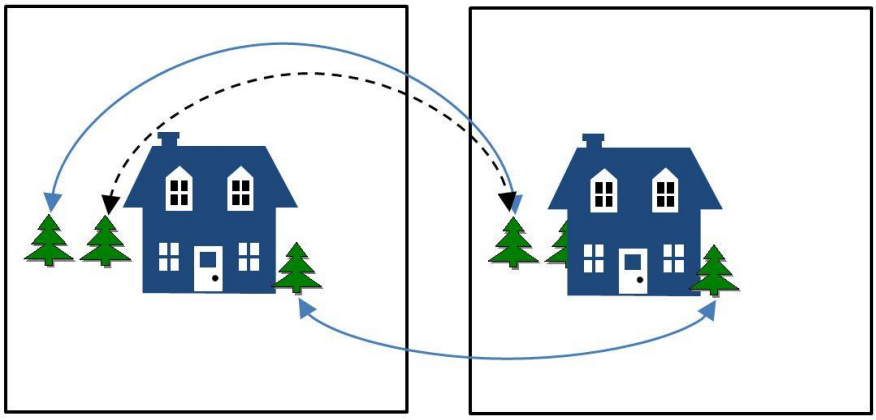


Figure 3-6 Bidirectional constraint applied to points in the background of a scene.

Many algorithms include an additional post processing step for improving the disparity estimate to subpixel accuracy (Tian & Huhns, 1986). If the SSD similarity measure is used, the cost values near an optimum can be approximated by a second-degree polynomial. Given the cost values of the optimum and its two nearest

neighbours, the subpixel estimate can be computed by the following equation (Mühlmann et al.,2002):

$$d_{subpixel} = d + \frac{lowest_{d-1} - lowest_{d+1}}{2(lowest_{d-1} - 2lowest\_value + lowest_{d+1})}, \quad (3.3)$$

where $lowest_{d-1}$ and $lowest_{d+1}$ are the two neighbours values in the dissimilarity function to the $lowest\_value$.

A step termed disparity calibration based is used as post processing in (Gu et al., 2008).The step is based on the assumption that in a finite region points which have similar colour and short spatial distance should have the similar depth. The disparity distribution of all pixels within a window following the above assumption is calculated, and the disparity which has the highest frequency of occurrences is assigned to the interest point as the final disparity. This step can be viewed as disparity smoothness by segmentation

3.5 Stereo Correspondence Evaluation Methodologies

The assessment of the progress on the stereo vision field is quite difficult if only qualitative results on the performance of algorithms are reported (Scharstein & Szeliski, 2002). Moreover, the lack of accurate and reliable disparity ground-truth information makes it quite difficult to compare different stereo correspondence methods even although results on the same imagery are provided (Hsieh et al., 1992).

A quantitative evaluation of disparity maps allows a comparison of stereo correspondence methods. A classification commonly found in the literature for quantitative evaluation approaches is based on classified into ground-truth-based and prediction-error based approaches (Szeliski & Zabih, 1999). Nevertheless, in practice, the most distinctive aspect of the evaluation process is based on what is used to compare against. Thus, the commonly used classification may be not suited to properly describe different approaches.

In this chapter, two main approaches are considered: evaluation approaches in the presence of disparity ground-truth data and evaluation approaches in the absence of disparity ground-truth data.

Evaluation methodologies in the presence of ground-truth data rely on measuring error by comparing an estimated disparity map against a ground-truth disparity map. A ground-truth disparity map for synthetic data can be generated by a raytracing algorithm, according to a model of the imaging environment on which the stereo method of interest is going to be deployed. Systentic data has been used in quantitative evaluation due to the difficulties to generate ground-truth on real imagery. However, synthetic data may fail to model the complexities of real-world, or in contrary, be artificially of a high complexity (Maimone & Shafer 1996; Scharstein & Szeliski, 2003). In fact, the generation of disparity ground-truth may be too difficult or laborious and even impossible to achieve in some circumstances due to the limitations of active stereo techniques to be used in indoor or controlled environments (Strecha et al., 2008, Morales & Klette, 2010). Ground-truth disparity maps on real imagery can be generated using an active stereo technique such as structured light (Scharstein & Szeliski, 2003), laser rangers (Mulligan et al., 2001) time of flight measurements, or even being manually generated (Hsieh et al., 1992) among others.

Most of the evaluation methodologies, in the absence of ground-truth, data rely on using additional views. Thus, the generation of additional views is a requirement that should be taken into account during the capturing process of the stereo image (Morales & Klette, 2009). In this type of approaches, a more complex camera setup or a more expensive camera system is required in order to fulfil this requirement.

3.5.1 Evaluation without Disparity Ground-truth Data

A comparison of four different stereo correspondence methods is conducted in (Bolles et al., 1993). Presented data show percentages of agreements on disparity estimation between estimated maps by considered methods. However, this approach is not capable of evaluating each method independently, neither capable of quantitatively characterising their performance (Maimone & Shaf, 1996).

The prediction-error methodology is proposed for the evaluation of both, motion and stereo correspondence algorithms in (Szeliski, 1999). A predicted view can be rendered based on a reference image and its associated estimated disparity map. Then, the rendered view is compared against an additional image (i.e. an image that was not used to compute the disparity map), captured from a known camera position with

respect to the input stereo image. Forward and/or inverse predictions are the two alternatives to generate a rendered view. However, error scores reflect not only the accuracy of the disparity estimation algorithm, but also the accuracy of the selected rendering algorithm, since the rendering process of the predicted view has to deal with interpolation or extrapolation issues (Scharstein & Szeliski, 2002; Sellent & Wingbermühle, 2012). In practice, this evaluation approach is best suited (or conceived) for applications domains where the output is a rendered view and human observers are final users. In this applications domain scenario, the capability of bringing a visual comfort sensations to a user could be more important that the accuracy of the estimation (Meesters et al., 2004). In the evaluation presented by the author, the goal was to conduct an inter-technique evaluation. The experimental evaluation conducted in this work focuses more in motion estimation, than in disparity estimation. In this methodology, the Root Mean Square (RMS) is used as error measure.

A methodology termed as self-consistency is presented in (Leclerc et al., 2000). This methodology is motivated on a property of the Human Visual System (HVS), on which perceptual inferences made by a HVS, from different viewpoints, are, most times, consistent among them. The self-consistency property is used for assessing the performance of disparity estimation algorithms, by measuring the distance among triangulated 3D world coordinates of a set of corresponding points from multiple views. Nevertheless, there are two main issued regarding the self-consistency property. On the one hand, the assessment of disparity estimation algorithms by the self-consistency property requires reliable information about projection matrices, in a common coordinate system. In fact, the knowledge of information can be considered in practice as having ground-truth data. On the other hand, the self-consistency is a necessary but not a sufficient condition for a disparity estimation algorithm to be correct. Consequently, an algorithm can be self-consistent over several scenes, but it may produce severely biased or entirely wrong disparity estimations. Thus, although an evaluation based on the self-consistency property is quantitative, it can be considered as heuristic. The authors use their methodology to conduct intra and inter-technique comparisons.

The use of the prediction error approach for stereo sequences in the context of vision-based driver assistance systems is discussed in (Morales & Klette, 2009). The use of sequences allows to observe and analysing the impact of varying conditions

within the domain context. The used camera setup involves three calibrated cameras in a vehicle capturing real data. The cameras at the centre and at the right are generating the reference and the target views, respectively. The third camera is collinearly located at the left of the reference camera. The conducted evaluation also includes synthetic data on which the third image acts as ground-truth. Global stereo methods based on dynamic programming, belief propagation, semi global matching (Hirschmuller, 2005), and graph cuts were selected for comparison. The RMS and the NCC were used as evaluation measures for comparing the virtual view generated based on estimated disparities against the third view, along the considered stereo sequences. It is highlighted that different results were obtained on synthetic data being the semi global and the graph cuts based methods the top performer according to RMS and NCC measures, respectively. A similar situation arise on real data, being the dynamic programming and the belief propagation methods, the top performer according to RMS and NCC, respectively. Authors point out that the NCC measure is more suited than RMS to be used in real imagery, and the difference between evaluation results obtained using synthetic data against real data may be due to the lack of challenging varying image content situation in the former data.

3.5.2 Evaluation Based on Disparity Ground-truth Data

The results of several stereo methods are compared in (Guelch, 1991) using a synthetic image, and manually created disparity ground-truth data. Eleven pairs of images of 240 x 240 pixels were matched. The work was motivated as an effort to determine the state-of-the-art in stereo correspondence methods. It aims to address problem such as: how stereo methods behaves on images of different complexity, how much a priori knowledge is required by the methods, and how far the methods assess the quality of estimation, as well as the accuracy of obtained results. The standard deviation was used as measure for evaluating estimated disparity maps. The quality assessment was performed by several participants belonging to different institutions around the world. It was done either automatically, or manually, or in a combination of both.

An evaluation and comparison of stereo correspondence methods of different approaches is presented in (Hsieh et al., 1992). Although the evaluation is motivated for

the cartography application domain it is quite complete and suitable to be used in other domains. The paper address the matching of aerial images using a feature based and an area based methods, as well as the combination of the results produced by both methods. The authors developed a display tool to assist a user in the generation of disparity maps. A user selects and matches a sufficient quantity of points, which are later used in the generation of a dense disparity map by interpolation. The use of these maps allows the ground-truth based evaluation proposed. With regard to image content, evaluations of the entire image, of all the buildings, as well as in building-by-building basis are identified as interesting and required to be considered during the calculation of performance measures. Moreover, an evaluation based on the 3D properties of the captured scene, which are reflected on the disparity ground-truth data (i.e. regions of homogenous disparity or with disparity jumps) is also suggested. This consideration is used for evaluating how well the algorithms behave on depth discontinuities (i.e. building boundaries). In addition two measures for comparing estimated maps against ground-truth data are used (i.e. although not explicitly formulated neither named). These measures are the Mean of Absolute Differences (MAD) and the Good Matched Pixels (GMP) percentage, respectively. The GMP is formulated as follows:

$$GMP = \frac{1}{N}\sum_{(x,y)}^{N} \gamma\,(x,y)\,;\ \gamma(x,y) = \begin{cases} 1, & if\ (|D_{true}\,(x,y) - D_{estimated}\,(x,y)| \le 1) \\ 0, & otherwise \end{cases}, \tag{3.4}$$

where $D_{true}\,(x,y)$ and $D_{estimated}\,(x,y)$ are the ground-truth disparity value, and the estimated disparity, respectively at the position (x,y), and the value of 1 pixel is motivated based on the inherent error that may be present in the disparity ground-truth data, due to its construction process. In addition, the values obtained by the proposed measures for the evaluated maps are plotted according to the disparity range, as it is illustrated in Figure 3-7.

A distinction among disparity estimation errors based on the error magnitude is used in (Lan et al., 1995). In this way, errors up to one pixel are considered as good matches, errors up to two pixels are considered as acceptable matches, errors up to three pixels are considered as failed matches, and errors of more than three pixels are considered as false matches. In addition, the concept of evaluation criteria is used

empirically for analysing errors in a small portion of the image near to depth discontinuity and far from such region.

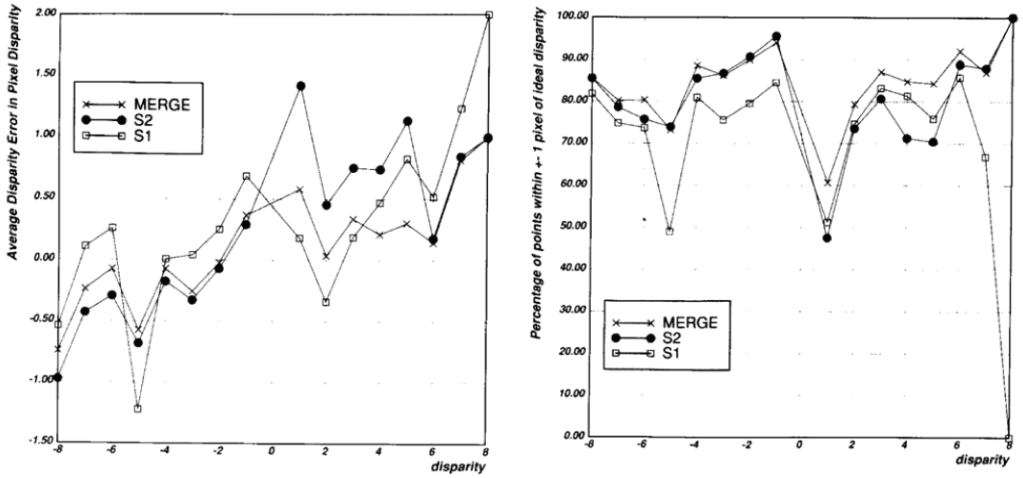


Figure 3-7 Illustration of average error and percentage of errors, respectively, according to disparity values (Hsieh et al., 1992).

A taxonomy for stereo vision computer experiments is proposed in (Maimone & Shafer, 1996). This work highlights the relevance of having disparity ground-truth data in order to properly measure by how much estimated disparity maps differ from the truth. Authors point out the convenience of representing stereo data, including the ground-truth data, by disparity maps against depth maps and object models, since it allows different stereo methods to be compared on an equal basis. Moreover, disparity data can be converted into depth by including the parameters of the actual camera system in the data set. The use of occlusion masks, represented by binary images, is suggested in order to allow an evaluation only on stereo visible points. The comparison of estimated disparity against ground-truth considers a tolerance threshold. They outline a taxonomy of ground truth-scenarios including noiseless synthetic data, noisy synthetic data, controlled environment, a measured environment, and an unconstrained environment.

An experimental comparison of stereo correspondence algorithms is proposed in (Szeliski & Zabih, 2000). This proposal applies, in a separate way, a comparison against disparity ground-truth data, and the prediction-error approach of (Szeliski, 1999). The concept of error criteria is introduced in this proposal. The use of error criteria allows a detailed analysis of algorithms performance in relation to different image phenomena, such as specular surfaces, low texture regions, depth discontinuities, and occluded

pixels, among others. In addition, an error function for disparity maps evaluation is also introduced. It defines an error as an estimation disagreeing from the ground-truth disparity value in more than a threshold. The error function is gathered according to error criteria. The Tsukuba and the Map stereo images were used as test-bed images. The ground-truth of the Tsukuba image was generated manually, which make it prune to errors, whilst the Map image is, in essence, an artificial image. On the other hand, the prediction error approach is used considering how well the reference image and its estimated disparity map can be used to predict other views using a forward or inverse warping. The authors conclude that consistent results were obtained, whilst each approach detects better particular kinds of errors: the prediction-error based approach does emphasis on errors over highly textured regions, and the ground-truth based approach does emphasis on errors over low texture regions.

A performance analysis of stereo methods with respect to the task of immersive tele-presence visualisation environment is presented in (Mulligan et al., 2001). In the tele-presence environment, a scene is displayed stereoscopically and the scene changes according to the point of view of the user. An evaluation test-bed, providing a set of depth ground-truth data, involving multiple views of the face of mannequin and captured by laser measurements is provided. It involves three metrics which are closely related to the application domain (i.e. the view independent world-centred depth difference, and differences between novel rendered views and projected ground-truth views) regarding the quality of experience. This type of evaluation keeps similarity with the evaluation proposed in (Barron et al., 1994, Szeliski, 1999). An experimental study of the effects on occlusion and low texture on the distribution of error metrics is provided. The evaluation is used to compare the performance of two (a local and a global) area based algorithms, in the context of the considered application domain.

The Middlebury's methodology was presented in (Scharstein & Szeliski, 2002). This methodology extends concepts previously introduced in (Szeliski & Zabih, 1999). It can be used in both intra and inter-techniques evaluation. Evaluation processes conducted following this methodology allows comparing estimated disparity maps by different algorithms. The imagery test-bed includes four images, and their respective ground-truth data: Tsukuba, Map, Venus, and Sawtooth. The error function used in

(Szeliski & Zabih, 2000) is properly formulated and termed as Bad Matched Pixels (BMP).

$$BMP = \frac{1}{N}\sum_{(x,y)}^{N} \varepsilon\,(x,y)\ ; \varepsilon\,(x,y) = \begin{cases} 0, & if\ (|D_{true}\,(x,y) - D_{estimated}\,(x,y)| \le \delta) \\ 1, & if\ (|D_{true}\,(x,y) - D_{estimated}\,(x,y)| > \delta \end{cases}, \quad (3.5)$$

In addition to the BMP, the RMS is also used for comparing estimated maps against ground truth data. Different error criteria are associated to image segments resulting in the following criteria:

- *all*: the entire image,
- *nonocc*: areas that are not-occluded,
- *disc*: areas near depth discontinuities and occluded regions, and
- *textureless*: areas of low texture.

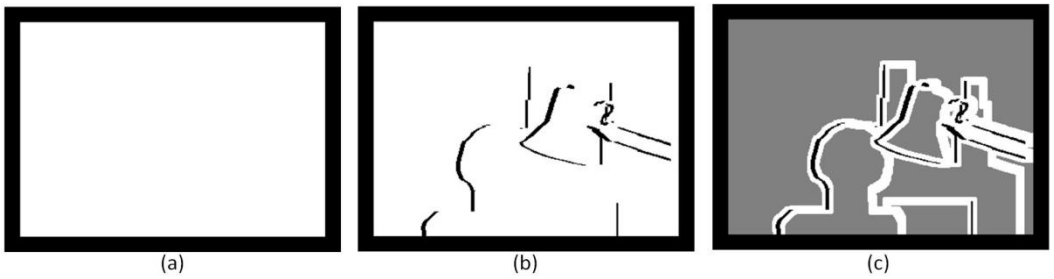


Figure 3-8 Masks associated to evaluation criteria of the Tsukuba image: (a) *all*, (b) *nonocc*, and (c) *disc.*

In the evaluation model of this methodology, error scores are sorted in descendent way by each error criterion and stereo image pair. A final rank is computed as the average of all ranks. In this way, the evaluation model of Middlebury's methodology can be seen as a linear combination of ranks, where a real value is associated to the accuracy of an algorithm. Different experiments were conducted and obtained results were plotted for analysis. The authors conclude that the conducted experimentation produces a better understanding on shortcomings of some algorithms with regard to their particular composition in terms of constitutive modules or building blocks, as well on the sensitivity of algorithms to setup of key parameters.

The performance of a small group of stereo methods in noised synthetic images is analysed in (Leclercq et al., 2003). A stereo method based on dynamic programming (Birchfield & Tomasi, 1998) as well as a local based method of fixed support area, using different matching costs (i.e. SSD, SAD, NCC, Census transform) were considered during the comparison. Disparity maps were compared against ground-truth data using three evaluation measures: the fraction of correctly computed disparities (i.e. percentage of estimated disparities within 0.5 pixels of the true disparity), the mean and standard deviation of the distribution of disparity errors. The execution time was also compared but in a separated way. The authors pointed out the first used metric as the most important one, and the other two as complementary. Different levels of additive white Gaussian noise were independently added to the RGB channels of the two colour stereo images generated by a ray tracing algorithm. The inter-technique evaluation of the local method is focused on the impact of varying the window's size, whilst the inter-technique comparison of the disparity based method was focused on tuning the terms related to the reward for a match and the penalty for occlusions. The best found combination of parameters in both methods was used to evaluate the degradation of performance in the presence of noise. However, it was found that in the case of the dynamic programming based method the parameters obtaining the best results in noise presence do not coincide with the parameters obtaining the best results without noise. Authors conclude that, among the considered methods, the dynamic programming showed the better results.

An online rank based on the Middlebury's methodology is available and keeps updated at (Scharstein & Szeliski, 2012). The imagery test-bed used is composed by the Tsukuba, the Venus, the Cones and the Teddy stereo image pairs (Scharstein & Szeliski, 2003). The BMP and the *all*, the *nonocc* and the *disc* are used as error measure and error criteria, respectively. The evaluation is focused on an inter-technique comparison, and algorithm's parameters have to be fixed for the entire test-bed. This online benchmarking has been widely used by the community. It contains a repository of disparity maps generated by, approximately, 140 algorithms (November, 2012). Nevertheless, it compares the entire set of algorithms reported by the community regardless their differences in requirements (e.g. used hardware, or execution time, among others). Thus, such comparison may be unfair.

An evaluation of real-time stereo correspondence methods, in the context of autonomous vehicles, is presented in (Mark & Gavrila, 2006). The evaluation involves real imagery without disparity ground-truth data, and synthetic imagery with ground-truth. The synthetic imagery is a stereo sequence recreating an urban environment with traffic conditions. Perturbations related to real imagery capturing conditions were added to synthetic imagery. In this way, obtained results on both types of images were consistent. Image points are classified into background obstacles, road surface and sky, in order to gather estimation errors only in relevant points within the application domain. Authors discussed about the properties of several evaluation measures such as the MAE, the SAE, the BMP and the MRE, pointing that only the MRE is able to consider the inverse relation between depth and disparity. The average of the percentage of points invalidated by left-to-right checking, estimation density and MRE, over multiple frames are used as evaluation measures. Authors pointed out the relevance of incorporating into the evaluation results of subsequent steps such as obstacle detection.

The performance of stereo methods for 3D face reconstruction is assessed in (woodwar et al., 2006). Disparity ground-truth data is created using a coloured grid projected onto a human test subject. A stereo pair of images, as well as a third one for texture mapping purposes are captured with and without grid projections. The correspondence between left and right images is manually conducted at grid intersection points. The sparse dataset is transformed into a dense one by cubic spline interpolation and 3D face model fitting. The methods selected for comparison were: local methods using different matching costs (i.e. SSD, SAD, NCC, Census transform) over a fixed support region, and global methods based on seed propagation (Chen & Medioni, 1998), Markov chain modelling (Gimel´farb, 2001), and different graph cuts based methods (Roy & Cox, 1998; Boykov et al., 1999; Kolmogorov & Zabih, 2002). This work uses the same evaluation measures used in (Leclercq et al., 2003). Authors highlight that the local method using the Census transform showed the highest fraction of correct disparities, but other methods with a lower fraction of correct disparities showed a lower standard deviation than the obtained by using Census based method. Authors conclude that the method using the SAD matching cost obtained best results in a low execution time. This fact may be related to the image content which is mainly dominated by a face (i.e. which is indeed a smooth surface), as the foreground plane, and the background, on

which occlusions may arise but it has no relevance within the particular application domain.

An evaluation conducted on a real-time stereo platform, based on the processing power of programmable graphics hardware is presented in (Wang et al., 2006a). Different cost aggregation approaches are implemented and optimised for graphics hardware in order to maintain real-time speed performance. In some cases, such as for the proposal of (Yoon & Kweon 2005), it requires a simplification of the matching cost. They are compared, under a WTA optimisation strategy, in terms of both the computation time required and the quality of the disparity maps generated, but in a separated way. The quality of the estimated disparity maps is assessed using the Middlebury's evaluation methodology (Scharstein & Szeliski, 2002). The performance is compared in terms of execution time. An extension of this work is presented in (Gong et al., 2007). A more detailed discussion on the implementation of different cost aggregation approaches is provided. Again, the Middlebury's evaluation methodology is used to compare the accuracy of estimated disparity maps.

A comparison of local methods in the context of automated biological control for agriculture technologies is presented in (Nielsen et al., 2007). The image content in this context is challenging due to the non-rigid structure of plants, which can be very complex, non-planar, and varying in orientation and height. Synthetic images generated using a ray trace algorithm, and considering image capturing problems (i.e. presence of Gaussian noise, focal blur, imperfect epipolar alignment, and imperfect brightness/contrast between cameras) were used were used in the comparison. The BMP was used as evaluation measure with a threshold computed based on the MSE. In addition, images of real plants (mostly containing just one plant per image, with a well differentiated background) with similar structural categories were annotated for comparison in order to validate the performance results obtained in synthesised images. Authors conclude that obtained results on real imagery are comparable with those obtained using synthetic images simulating capturing problems.

The evaluation methodology proposed in (Kostlivá et al., 2007) is focused on parameter settings (i.e. it addresses an inter-technique evaluation scenario prior to an intra-technique evaluation). It considers both the accuracy and the density of the

estimated disparity map with regard to parameters settings. Two errors are defined based on the accuracy and the density of estimated disparity maps: the error rate and the sparsity rate. The error rate is defined as the percentage of incorrectly estimated disparities, without considering a missing disparity as an error (i.e. mismatches and false positives). The sparsity rate is defined as the percentage of all missing disparity estimations which are not ruled out by any other incorrect estimation (i.e. false negatives). These errors are based on four principles: orthogonality, symmetry, completeness and algorithm independence (Kostlivá et al., 2003). These principles can be outlined as follows.

- Orthogonality: Errors definitions have to be mutually independent.
- Symmetry: errors have to be invariant to the direction of search (i.e. from the left to the right view, or from the right to the left view) during the disparity estimation process.
- Completeness: error definitions have to be valid in any scene of arbitrary geometry.
- Algorithm independence: an evaluation process has to be possible, disregarding the density, or semi-density, of estimated disparity maps.

A Receiver Operating Characteristics (ROC) analysis is adopted upon the error rate and the sparsity rate. In addition, an "*is better"* relation is defined based on the ROC curve. A particular parameter setting is better than another if it produces more accurate and denser results. However, the ROC curve can be computed on just one set of stereo images. Thus, the evaluation may turns probabilistic when the imagery test-bed includes several stereo image pairs. Additionally, this evaluation methodology requires weight assigns in relation to the importance of each stereo image pair included in the test-bed. Moreover, although the algorithm independence is on the principles motivating the methodology, it assumes that should be there different evaluation models depending on the density or semi-density of estimated disparity maps. Two stereo images, of artificial scenes with varying texture, were used in the experimental evaluation in addition to the Middlebury's imagery test-bed. The authors conclude that, disparity estimation algorithms with different occlusion models should not be compared, the criteria for

selecting test-bed images is still an open debate, and algorithms execution time should be also considered during an evaluation process.

A cluster ranking intra-technique evaluation method is proposed in (Neilson & Yang, 2008). The authors point out that most of evaluation approaches only consider a test-bed imagery of a very small size, and consequently, the obtained results may lack of statistical significance. It is assumed that evaluation results should be of general character (i.e. to be repeated for any other imagery test-bed). The proposed method consists on using a statistical inference technique (ANOVA) to rank the accuracy of disparity estimation algorithms over a single stereo image pair, and the posterior combination of ranks from multiple stereo pairs, into a final rank. Thus, the same ranks are assigned to algorithms producing statistically similar results. This proposal is focused on comparing matching costs using a hierarchical belief propagation algorithm (Felzenszwalb & Huttenlocher, 2004). However, a different significance test (Friedman) has to be applied when the test-bed includes more than one stereo image pair. Moreover, a greedy clustering algorithm, which requires a threshold related to a confidence level, is used. The clustering algorithm used, computes iteratively the final ranks as the average of several ranks in a partition. Thus, assigned rank may be a real number which lacks of a concise interpretation. This conducted evaluation includes 90 synthetic images, with three different levels of noise, generated by a ray tracing method, and 18 images from the Middlebury's image repository, some of them captured with three different illuminations and times exposure (Hirschmüller & Scharstein, 2009). The BMP measure is used, only, according to the *nonocc* error criterion. The authors conclude that the selection of a matching cost has a large impact on the accuracy of estimated disparity maps, and there is not a single parameters setting working well for every matching cost metric or even every stereo image pair. Moreover, a particular setting, working fine in one case, may worked very poorly in other case. On the other hand, this class of study requires huge computational resources, which are not available for all developers or researchers. Thus, the conducted study is difficult to be repeated. Consequently, this methodology may not take into account the capabilities or requirements of final users.

A framework for evaluating short-baseline stereo-based pedestrian detection techniques is proposed in (Kelly & O'Connor, 2008). The work is motivated for the

advantages of stereo-based pedestrian techniques over conventional 2D-based pedestrian techniques (Zhao & Thrope, 2000; Kelly, 2007), as well as by the lack of a standard stereo data set suited for the application domain, and an agreed methodology for carrying out the evaluation. Synthetic data with disparity ground-truth and real imagery with specific content regarding the addressed application domain are provided. Generated data aims to incorporate the challenges that may arise in a real world scenario. A set of evaluation metrics are recommended, and two evaluation methodologies for the specific domain are proposed: the first is based on traditional image plane comparison techniques, and the former is based on 3D properties. The proposed framework uses the Middelbury's methodology for evaluating the accuracy of the disparity estimation process by the stereo correspondence methods applied on generated data. The average of obtained RMS scores was used in order to determine what error tolerance should be used for the in the BMP measure. These values were compared against the values obtained by a stereo correspondence method specifically designed for the pedestrian detection domain.

In (Vaudrey et al., 2008) it is pointed out that the test-bed imagery considered for evaluating stereo (and optical flow) should not be only synthetic (i.e. generated by a ray tracing or captured under engineered conditions), since conflicting results may be found when real imagery is used: improvements made on synthetic imagery are not always translated to improvements on real images, as well as methods working well on real imagery, may do not perform well on synthetic data, and vice-versa. Moreover, it can be differences in results due to the quantity of bits and image type between synthetic imagery (8 bits colour images) vs. real imagery captured by industrial cameras (10 to 12 bits greyscale). The first two sets of the EISATS dataset are presented in this work. These sets includes seven real imagery stereo sequences of driving scenes with different manoeuvres in a relative low traffic conditions in different environments (i.e. highway and semi-urban areas), and illumination changes, captured by calibrated industrial cameras, as well as synthetic long stereo sequence with ground-truth data. It is pointed out that in not all application domains the accuracy of estimation is the most important factor to be evaluated such as in driver assistance systems, for instance, on which the robustness has a higher priority than absolute precision. The presented dataset is used to analyse the behaviour of methods performing well on other datasets

(Scharstein & Szeliski, 2012). It was found in some cases that a pre-processing for real imagery (i.e. compute the gradient) improves accuracy of disparity estimation, but, such pre-processing applied to synthetic imagery may decrease (although slightly) accuracy on obtained results. Some possible reasons are identified for the discrepancies between results on synthetic against real imagery: poorly defined objects boundaries and radiometric differences between stereo images.

A performance evaluation scheme and metrics for stereo methods at three different levels in the domain of automotive applications are introduced in (Steingrube et al., 2009). The first level (i.e. low-level) includes evaluation at pixel level using knowledge about object-free in order to detect mismatches. The second level (i.e. mid-level) evaluates the disparity data roughly column by column in the object-free space at the front of the car. The third level (i.e. high-level) performs an evaluation on an object level based on leader vehicle measurement. In particular, mismatches are detected based on assumptions about an object free-volume in the 3D scene in front of the vehicle, where the ground-truth data is implicitly given by the assumption that the road is planar and the moving vehicle does not collide with any object in a specific time window. Thus, if a stereo method computes a 3D point within this volume, a false correspondence has been found. They are evaluated based on the ratio between false correspondences and the total of computed correspondences estimated by a method. Nevertheless, this metric may be sensitive to the density of the estimation. It does not consider the location of the false correspondence within the object-free space, but this aspect is cover by the med-level evaluation. Metrics are integrated in a frame basis to provide a single score for the whole sequence. Presented three level analysis aims to cover the range of applications in which stereo methods are most used in automotive industry. A large amount of test-bed imagery of uncontrolled scenes and different weather conditions are considered for evaluation purposes. Ground-truth for this data is semi-automatically generated: starting with a well performing stereo method (i.e. at least qualitatively judged) an initial ground-truth is generated, and it is iteratively improved by human supervision over additional estimations. Three real-time stereo methods are compared, (i.e. one local and two global) and a semi-global matching based method shown best evaluation results in all considered metrics. A similar proposal to this work is presented in (Schneider et al., 2011). It includes an evaluation in compact medium-level

representation describing local three dimensional environments termed Stixel World (Pfeiffer & Franke, 2010), as well as a more detailed description of the three levels of evaluation used.

An inter-technique evaluation, involving estimation accuracy and computational efficiency, is proposed in (Tombari et al., 2010a). This proposal is focused on disparity estimation algorithms suitable to be used in application domains requiring near real-time performance (i.e. more than 5 frame per second), real-time performance (i.e. more than 25 frame per second) and/or to be executed on hardware platforms with limited resources (i.e. with a low-power architecture and limited memory) such as the offered by embedded devices. The imagery test-bed used includes the Middlebury's data set (with the addition of Gaussian noise) as well as a dataset related to common working conditions (i.e. uncontrolled illumination, photometric distortions, small defocus, and non-perfect rectification). This dataset, termed Lab, was acquired in an uncontrolled environment using an off-the-shelf stereo camera. It is composed by 6 stereo images containing different objects on a table (Tombari et al., 2010b). The complement of the BMP measure was used to gathering errors according to the *nonocc* and the disc criteria, whilst the computational efficiency was compared based on the quantity of disparities computed per second. This proposal uses the evaluation model of Middlebury's methodology. Nevertheless, accuracy and efficiency are, by nature, conflicting goals for a disparity estimation algorithm. The authors conclude that, in top performer algorithms, a small increment in estimation accuracy implies a large additional computational effort. In addition, the Lab dataset may be much more challenging than the Middlebury dataset.

In (Haeusler & Klette, 2010) is pointed out that it might be possible to quantify quality of recorded stereo images with respect to some measures, which may be used for indicating domain of relevant scenarios when performing evaluations for some particular test data. The aim of the work is to judge the complexity of a specific stereo dataset and its qualitative relation to other datasets. In particular, the SIFT-descriptor (Lowe, 2004) is applied on rectified images, but without the epipolar constraint. It is observed that matches obtained by the SIFT-descriptor in synthetic or engineered images are mainly same-row matches. Two measures based on SIFT matching counts are proposed: the matching rate and the mismatch rate. The matching rate indicates

how many features on average lead to one match disregarding if it is correct or not, whilst the mismatch rate identifies the percentage of incorrect matches. A correct match is determined using a tolerance threshold of 1 pixel. It is pointed out that, according to proposed measures, outdoor stereo data captured by industrial cameras shown the highest level of complexity among the considered datasets, whilst the Middlebury's benchmark data set evaluates similar to synthetic stereo images of medium complexity. Among the presented conclusions, it is stated that synthetic data will remain important for testing stereo matching, especially due to having full control about the image formation process.

The Leuven dataset (Leibe et al., 2007), a stereo video shot captured from a moving vehicle drive in an urban environment, is augmented in (Ladicky et al., 2010) by incorporating manually estimated semi-dense disparity ground-truth data. The augmented dataset also contains object labels (i.e. road, building, car, sky, person, bike, sidewalk, and void for points related to none of the above). The augmentation of the dataset data was performed in two steps. Firstly, object labels were manually assigned. Secondly, semi-dense disparity maps were created by manually matching corresponding planar polygons. The imagery dataset contains large regions of homogeneous colour and texture. It also presents poor photo-consistency and inconsistent luminance between stereo views, as well as specular reflections. With regard to evaluation measures, the BMP is used under different thresholds.

Stereo methods are compared against ground-truth data generated by a laser range finder in the context of vision-based driver assistance systems in (Morales & Klette, 2011). The generated disparity ground-truth is sparse, with sub-pixel precision, but uniformly distributed in a depth field of interest from 5 meters to 120 meters. In the points where ground- truth is available, (which are less than a 10% of the size of the captured image) the evaluation is based on a commonly used evaluation measure (i.e. the BMP with a threshold of 1 pixel) whilst in points without ground-truth measure, matching confidence measures are used (Egnal et al., 2004). Given three near points in the disparity ground-truth data to the point being evaluated, two 3D patches are generated. The confidence measure is computed based on the Euclidean distance between the centroids of each patch and their respective deviation. Stereo correspondence methods based on dynamic programming, belief propagation, graph

cuts and semi global matching were selected for evaluation. Three *stop and go* stereo sequences without moving objects were captured and used during the evaluation. It was observed that the behaviour of considered method may vary along the sequence, according to the image content. In addition, used evaluation measures shown that a stereo method may have a lower number of mismatches, but these mismatches may be of a larger impact. According to obtained results, the dynamic programming based method show better results in the considered stereo sequence. However, obtained scores were averaged over the sequence. This may make more difficult a comparison and interpretation of obtained results.

The KITTI vision dataset is proposed in (Geiger et al., 2012). It covers the following visual tasks: stereo, optical flow, visual odometry/SLAM and 3D object detection. The capturing platform is composed by four high resolution cameras, (i.e. two grayscale and two colour) a laser scanner, and a localisation system, calibrated and synchronised among them, providing a semi-dense ground-truth. The dataset comprises 389 stereo and optical flow image pairs (i.e. 194 for training purposes and 195 for testing), stereo visual odometry sequences of approximately 40 kilometres length, and more than 200000 objects annotations in a cluttered and dynamic scenario captured by driving around a mid-size city in both rural and highways areas. Different challenges for stereo methods such as non-lambertian surfaces, large displacements, large variety of materials, as well as different lighting conditions were captured. The online stereo ground-truth benchmark show results for the first 20 grayscale images where the environment is static and aiming diversity among them. Two evaluation criteria are used: non-occluded pixels as well as all pixels for which ground-truth is available. The percentage of bad matched pixels using different thresholds (i.e. { 2, …, 5}) is used as evaluation measure for disparity maps, being 3 pixels the default value for which results are reported. This threshold value takes into consideration almost all calibration and laser measurement errors. Global stereo methods based on graph cuts (Kolmogorov & Zabih, 2001), semi global (Hirschmuller, 2005), variational model (Ranftl et al., 2012) Markov random fields (Yamaguchi et al., 2012), as well as seed growing (Cech et al., 2011) and local based methods were initially selected for evaluation. Missing disparities are filled-in for evaluated disparity maps using background interpolation (Hirschmuller, 2003). Authors point out that stereo methods ranking high in others benchmarks such as

in the Middlebury's, are performing below average when are tested on captured data. This may be due to the violation of assumptions made by some methods about image content. The original density of each map is also reported. The benchmark is available online (KITTI, 2012) and a MATLAB/C++ development kit is also provided. Authors conclude that the variational based method presented in (Yamaguchi et al., 2012) is showing the best results.

In (Sellent & Wingbermühle, 2012) it is pointed out that conventional evaluation approaches for stereo methods evaluates dense and non-dense estimation in the same way, ignoring the risks associated of sparse image disparity estimations entirely missing objects. Thus, a correct evaluation and comparison of non-dense image matches an evaluation should take into account the sparseness of a disparity field as well as the distribution of matches over the objects in the scene. In this work, a normalised histogram based evaluation (for stereo and optical flow) is proposed. It can handle both dense and not dense disparity maps. If estimated disparities are accurate, the normalised histograms of disparity map and disparity ground-truth data are similar in shape and amplitude. The Earth Mover's Distance is used for comparing histograms. (Rubner et al., 1998). The proposed evaluation is focused in two aspects: distribution and outliers. The former aspect evaluates how well matches are distributed among all different scene entities, whilst the latter aspect evaluates the presence and frequency of outliers. Nevertheless, the property to distinguish between different regions of error is loosed by building histograms overs the entire image. Moreover, due to the loss of locality, the histogram-based metric cannot distinguish between noise and reliable estimates. Consequently a random disparity map may be evaluated as similar to disparity ground-truth data. On the other hand, this evaluation approach is motivated by arguing that conventional evaluation approaches evaluate only where both estimated disparity and ground-truth data are defined. In practice, this is not true since, due to the use of evaluation criteria, the comparison against ground-truth is performed in a previously fixed set of points, and, consequently, a missing estimation on the disparity map is penalised by the evaluation measure. Otherwise, an empty estimated disparity map will report a perfect match. The authors end the paper by concluding that just not a single metric should be used for evaluating disparity maps.

With regard to alternative evaluation measures for comparing estimated maps against ground-truth, a modification of the Multi-scale Structural Similarity index (MS-SSIM) measure is introduced in (Malpica & Bovik, 2009). The proposed measure, termed R-SSIM, is capable of handling missing data in both, a disparity map under evaluation and used ground-truth data. As a conclusion, obtained results by the R-SSIM measure and obtained results by the BMP measure are statistically correlated. Nevertheless, the final ranking assigned to disparity estimation algorithms, using the evaluation model of the Middlebury methodology, varies considerably when the R-SSIM measure is used. In addition, the discussion about the analogy among the components of the MS-SSIM measure and the properties of a disparity map is not properly addressed. Consequently, the argumentation of why it is convenient to use the R-SSIM measure for evaluating disparity maps turns weak.

Two indices for measuring smoothness of a noised disparity map –the disparity gradient and the disparity acceleration– are proposed in (Zhang et al., 2009c). Three ground-truth disparity maps from the Middlebury repository (Scharstein & Szeliski, 2003) were artificially corrupted with noise. Then, the proposed indices were applied to the noisy images, and obtained results compared against the level of introduced noise. The indices require a threshold in order to consider an estimated disparity as inaccurate, or related to a depth discontinuity. However, no information is provided about how required thresholds can be fixed, neither about considered noise nor its relation to artefacts produced by a disparity estimation algorithm. Moreover, this work ignores the fact that a disparity map may vary smoothly, but being totally wrong. Consequently, the disparity gradient and the disparity acceleration indices may be not suited for properly evaluating an estimated disparity map.

The comparison of results using the SSIM and the PSNR measures on noisy DM by adding salt and pepper is addressed in (Shen et al., 2011). The authors conclude that obtained PSNR values are closer to the scores assigned by subjective evaluation. However, this conclusion does not coincide with the well-known drawback of the PSNR (Wang et al., 2004). Additionally, there is not a clear relation between the type and the level of noise introduced, and the artefacts that a disparity estimation algorithm may produce. Consequently, the considered evaluation scenario lacks of realism.

3.6 Decision Making in Multiobjective Optimisation Problems

As part of a decision making process, a solution or a very small set of solutions from the Pareto front, has to be selected in order to solve the problem being addressed. This selection is a responsibility of a decision maker. However, in most of cases, a Pareto front may overload the judging capabilities of a decision maker, due to factors such as its large cardinality (Ben Said et al., 2010), the multidimensional complexity of the problem being solved (Brockhoff et al., 2006), plus inherent limitations of a decision maker for effectively handling large amounts of data and more than several factors at once (Cvetkovic & Parmee, 2002), among others. Although a visualisation of the Pareto front may assist in the decision making process, a visualisation becomes complex with several solutions and three objectives or more, as well as visualised information for making a decision may become difficult to use (Das, 1999). Difficulties in a decision making process may be alleviated by introducing preferences (Rachmawati & Srinivasan, 2006). Preferences can be viewed as knowledge and/or expectations about a problem solution. They can be used as a mechanism to decide if a specific solution is preferable than other solutions (López & Coello, 2009). Nevertheless, in some cases a decision maker may lack of information for selecting a solution and/or has not preferences among all objectives. In the absence of preferences, it is generally assumed that the most preferable solution correspond to a region in the maximum convex bulge of a Pareto curve/surface, termed as the knee region (Das, 1999). However, identifying the knee region of a Pareto front requires solving a non-linear optimisation problem, as well as some a priori knowledge on a Pareto front. In addition, determining the knee region(s) may become prohibitively complex as the dimensionality of a problem increases (Rachmawati & Srinivasan, 2006).

In general terms, preferences can be specified by a decision maker in three ways: *a priori*, *interactive* and *a posteriori*. In the *a priori* way, preferences are specified before the beginning of search process by the aggregation of objective function into lexicographic order or into a linear/nonlinear combination, among others (Rachmawati & Srinivasan, 2006). A deep knowledge of the problem and a clear understanding of the search space are required.

In the *interactive* way, preferences are specified during the search, based on a progressively and interactively acquired knowledge of the problem (Ben Said et al., 2010). An intensive effort of a DM is required, since, he/she is asked to give preference information at each algorithm's iteration, commonly consisting in specifying aspiration levels for each objective function, classifying objective functions according to their relevance, or introducing references points, among others. However, a decision maker may have large optimistic or pessimistic aspiration levels. In addition, disagreement about preferences might arise among decision makers when there are several of them involved in the decision process. Preferences specified in a priori or interactive way have an impact on search results.

In the *a posteriori* way, the search is executed first, and after that, a decision method is applied into the Pareto front (Parreiras et al., 2006). In this case, a decision maker has too many choices to select from, and a fair comparison among them is not an easy task to achieve due to the inherent dimensional complexity. There are two main approaches, to perform a posteriori multi-criteria decision making: utility functions and outranking methods (Cvetkovic & Coello, 2005). Utility functions assign a numerical value to each solution. Outranking methods are based on pairwise comparisons of all solutions, in order to establish if there exists preference, indifference or incomparability. However, commonly used methods under these two approaches rely on weights that should be specified by a DM (Parreiras et al., 2006). Methods such as the average rank, the maximum rank, and the favour rank do not require weights (López & Coello, 2009). The average and the maximum rank can be seen as utility functions. The average rank uses multiple ranks considering each objective independently and a final rank is calculated as the average of previously assigned ranks, whilst the maximum rank takes the best rank as the global rank for each solution. In the favour rank, a solution x is preferred over a solution y, only if x outperforms y on more objectives than those on which y outperforms x. However, the maximum rank method tends to favour solutions with high performance in some of the objectives, but with a poor overall performance. In addition, the average rank and the favour rank may produce even ranks, or indifferences, respectively, very often. Moreover, none of them considers the magnitude on which a solution outperforms another according to the involved objective functions.

3.7 Chapter Summary

- An evaluation methodology is required for at least two reasons: to assert the capabilities of a stereo method in a particular evaluation scenario and thus estimate its effectiveness, as well as to provide a systematic way for evaluating (perhaps incremental) changes to stereo methods (Courtney et al, 1997).
- A variety of disparity ground-truth data, generated by different means, and associated to stereo images captured in different and challenging conditions is nowadays available in the literature. This fact increases both the relevance and the necessity of conducting a proper disparity ground-truth based evaluation process.
- Although different authors have used some measures considering the disparity estimation error magnitude, and pointed the importance of this matter, the BMP measure is perhaps the most widely used evaluation measure in the literature on stereo methods. Moreover, the inverse relation between depth and disparity is ignored on most of evaluation processes available in the literature.

CHAPTER 4.

EVALUATION OF DISPARITY MAPS

Chapter Contents

4.1 An Evaluation Methodology for Disparity Maps

An evaluation of estimated disparity maps allows an assessment and comparison of stereo correspondence methods. A fair evaluation process requires of a methodology. An identification of components of an evaluation methodology makes possible an analysis about possible weaknesses in methodologies available in literature. In fact, most of existing evaluation methodologies do not state explicitly the constitutive elements, methods and steps, used and followed, respectively. In the context of the thesis, a methodology is understood as a set of steps, methods, and elements. The methods are applied in a specific sequence of steps. A method is composed by a set of modules, transforming a given input into a well characterised output, and the elements are the resources used and/or produced by following the steps of the methodology. In this way, a quantitative evaluation methodology requires, at least, of the following elements:

- Test-bed imagery: it involves a set of two or more views capturing 3D scenes.
- Disparity ground-truth data: a set of reliable, accurate and complete (i.e. dense) as possible, measurements of the disparity of a 3D scene.
- Disparity maps: estimated maps by a stereo correspondence method.

- Evaluation criteria: a set of requirements to be considered during the execution of the evaluation methods, related to stereo image phenomena challenging stereo correspondence methods.
- Evaluation results: objective data about the accuracy of estimated disparity maps, making possible obtain conclusions about the behaviour of stereo correspondence methods.

In addition, it involves the following methods:

- Algorithmic components: a set of software modules or building blocks of stereo methods.
- Stereo correspondence methods: a set of algorithmic components interacting in a previously fixed and ordered sequence, taking a stereo image pair as input, and estimating a disparity map as output.
- Evaluation measures: a set of functions for comparing disparity maps against ground-truth data, or for comparing rendered views against real views, in the context of evaluation in the presence of ground-truth, and evaluation in the absence of ground-truth, respectively.
- Evaluation model: a set of functions and processes followed in order to obtain and interpret evaluation results.

The above elements and methods are interacting among each other, in an ordered sequence of steps, such as the one illustrated in Figure 4-1.

Without loss of generalisation, some evaluation elements and methods are formulated below, as an extension of the work presented in (Cabezas & Trujillo, 2011).

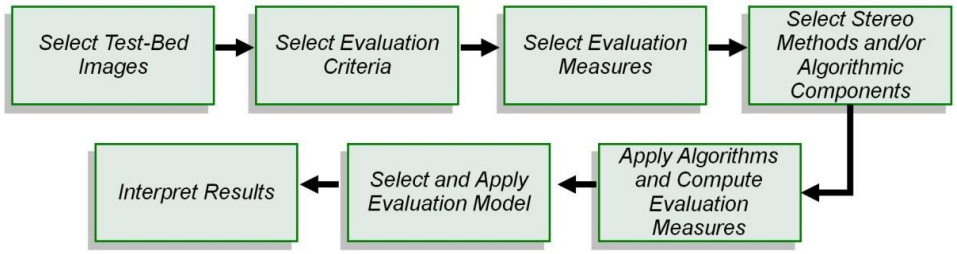


Figure 4-1 Steps for an Evaluation Methodology.

- Test-Bed Images

Let I_{stereo} be a set of stereo images:

$$I_{stereo} = \{I_1 , I_2 , \dots , I_k\}, \tag{4.1}$$

where I_k is the k^{th} view of a same 3D scene, subject to $k \geq 2$.

Let $I_{test-bed}$ be an imagery test-bed, as follows:

$$I_{test-bed} = \{I_{stereo_1} , I_{stereo_2} , \dots , I_{stereo_v}\}, \tag{4.2}$$

where I_{stereo_v} is a set of stereo images, and subject to $v \geq 1$, and to $\forall v$: $|I_{stereo_v}|$>2, in the case of evaluation in the absence of ground-truth.

Let D_{true} be a disparity ground-truth dataset:

$$D_{true} = \{D_1 , D_2 , \dots , D_v\} , \tag{4.3}$$

in the case of evaluation in the presence of ground-truth.

- Evaluation Criteria

Let $R_{criteria}$ be a set of evaluation criteria represented as image regions related to challenging surfaces and 3D properties, of captured scenes, for stereo correspondence methods:

$$R_{criteria} = \{R_{criteria_1} , R_{criteria_2} , \dots , R_{criteria_c}\}, \tag{4.4}$$

subject to $c = \{1,2, \dots , v\}$.

Let the function a be a stereo correspondence method:

$$a: \left(I_{test-bed_j}\right) \rightarrow M_{estimated(a)_j}, \tag{4.5}$$

where $M_{estimated}$ is a set of estimated disparity maps under evaluation:

$$M_{estimated} = \left\{M_{estimated(a)_j} \in M_{estimated} \mid \forall a \in A: \exists\, M_{estimated(a)_j}\right\}, \tag{4.6}$$

subject to $j = \{1,2, \dots , v\}$, and A is a set of stereo correspondence methods under evaluation:

$$A = \left\{a \in A \mid a: \left(I_{test-bed_j}\right) \rightarrow M_{estimated(a)_j}\right\}. \quad (4.7)$$

- Evaluation Measures

Let the function e be an evaluation measure producing a scalar value:

$$e: \left(M_{estimated_i} \times D_{true_i} \times R_{criteria_i}\right) \rightarrow \mathbb{R}, \quad (4.8)$$

to be used for evaluation in the presence of ground-truth data, and let the function q be an evaluation measure producing a scalar value:

$$q: \left(M_{estimated_i} \times I_{stereo_i} \times R_{criteria_i}\right) \rightarrow \mathbb{R}, \quad (4.9)$$

to be used for evaluation in the absence of ground-truth data, subject to $1 \leq i \leq j$.

Let $E_{eval(a)}$ be a vector of evaluation measures associated to the disparity maps estimated by the stereo correspondence method a, defined as follows:

$$E_{eval(a)} = \left\{e_{eval(a)_l} \in \mathbb{R}^v \mid e_{eval(a)_l} = e: \left(M_{estimated(a)_l} \times D_{true_l} \times R_{criteria_i}\right)\right\}, \quad (4.10)$$

subject to $l = \{1,2, \ldots, v\}$.

Let E_{eval} be a set of vectors of evaluation scores formulated as:

$$E_{eval} = \left\{E_{eval(a)} \in E_{eval} \mid \forall\, a \in A: \exists E_{eval(a)}\right\}. \quad (4.11)$$

4.2 A Review on Evaluation Methodologies Available in the Literature

A fair evaluation methodology should not introduce any type of distortion or bias into the evaluation process. In contrast to the intensive work addressing the stereo correspondence problem found in the literature, there are not so many works about issues related to the evaluation process looking to be of general purpose, or useful for the whole community researching on stereo vision. Moreover, although there is already an open debate regarding specific components of evaluation methodology (i.e. such as the nature and the quantity of involved test-bed imagery) there are some other

components which actual development still keeps resemblance with the first contributions on evaluation. Some of these components are identified below.

4.2.1 Imagery Test-bed

With regard to the test-bed imagery, the quantity of images that should be involved in an evaluation process is still an open debate (Trucco et al., 2013). Neilson & Yang (2008) pointed out that conclusions made upon evaluation results by considering only a few stereo images, may lack of statistical validity. Thus, a comparison of algorithms should involve a large quantity of stereo image pairs. Nevertheless, the required computational effort for processing such amount of data may turn an evaluation process of such scope beyond the capabilities of a final user. In addition, the image content has also to be considered (Vaudrey et al., 2008). On the one hand, it requires a trade-off between specific and general image contents according to the considered application domain. On the other hand, it has not been proved yet that an algorithm showing a good behaviour in a specific imagery test-bed will show also a good performance in a different test-bed. In practice, the lack of availability of a wide real imagery ground-truth dataset, suitable to be used in several application domains, and captured under non-controlled conditions imposes a problem for evaluating disparity maps and comparing stereo correspondence methods. In this work the selection of the Middleburry's imagery test bed was mainly motivated due to it's widely use by the stereo vision community. Although it is a small set, one of the aims of this work is to highlight the relevance of the other elements and methods involved in an evaluation process.

4.2.2 Evaluation Criteria

The use of evaluation criteria allows a detailed analysis on the behaviour of stereo methods on image challenging regions. This detailed analysis may provide information to a user about aspects on which the stereo method under evaluation is behaving well, or in contrary, about aspects on which it requires adjustments and or improvements. With the improvement on global stereo methods, as well as on adaptive local methods, an evaluation criterion such as the lack of texture on images started to fall into disuse (Scharstein & Szeliski, 2012). However, conventionally used criteria may be one step behind with regard to modern requirements such as detailed analysis on

occluded areas, among others. In addition, conventionally used criteria are simply assumed and used as binary segmentations o images, without a proper theoretical support. This is not a trivial issue since if a same image point is included in multiple binary segmentations, it will be taken into account more than once during the computation of evaluation measures, introducing biasing into obtained scores. This can be seen as implicitly introducing different weighting factors to points located in different image regions, without the user being aware of it. Moreover, these weighting factors are translated to the evaluation model. Consequently, the evaluation results may be biased, and difficult to interpret, even for a particular criterion, which was the original motivation. This problem is illustrated, in a general way, in Figure 4-2, using the distribution of *all* (points for which the disparity ground-truth value is available), *disc* (points near to depth discontinuities), and *nonocc* (non-occluded points) criteria (Scharstein & Szeliski, 2012). The consequence of this multiple inclusion is exemplified in terms of the cardinality of the masks vs. the image size shown in Table 4-1. It can be observed that, for the whole imagery test-bed, the quantity of evaluated points, obtained by summing the points considered for each criterion, exceeds the quantity of the points contained in the image. A specific case of this ambiguity is illustrated in Figure 4-3, using the evaluation criteria for the Teddy stereo image. It can be observed that the membership of a specific point to a particular criterion is not exclusive: points included in the disc criterion, are also included in the *nonocc* and the *all* criterion. In addition, the points in the *nonocc* criterion are also included in the *all* criterion. In fact, it seems to be contradictory that a same point is included in more than one criterion, since it is not clear which is the (most) challenging image phenomenon characterising it.

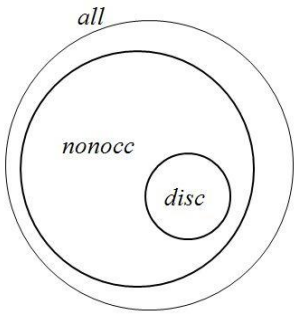


Figure 4-2 Relation among conventionally used error criteria.

Table 4-1 Ambiguous counting of error using conventional criteria.

	Tsukuba	Venus	Teddy	Cones
all	87696	150282	165344	163321
nonnoc	85438	147513	147651	143926
disc	15790	10540	40517	47189
Evaluated Points	188924	308335	353512	354436
Image Size	110592	166222	168750	168750

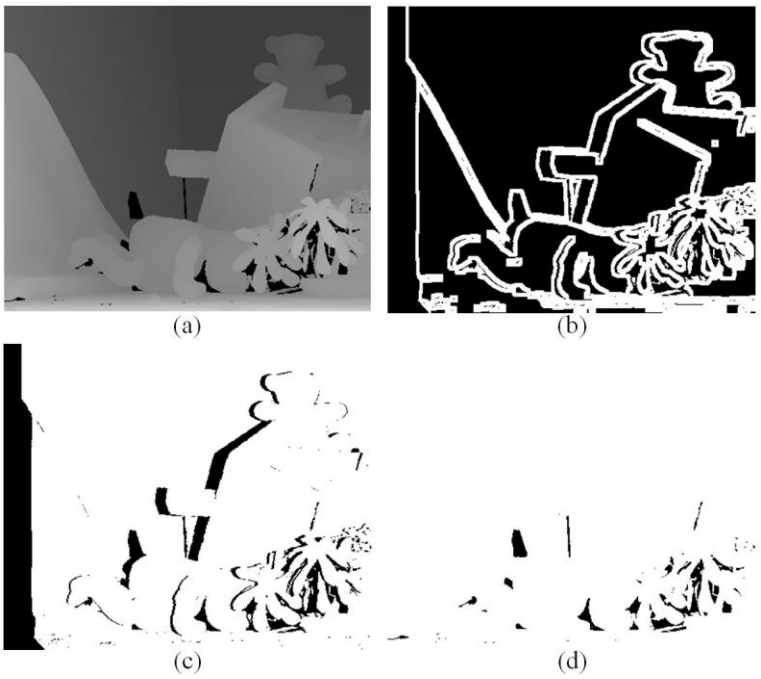


Figure 4-3 Conventional Evaluation criteria for Teddy image. (a) ground-thruth, (b) *disc* criterion (c) *nonnoc* criterion, and (d) *all* criterion

4.2.3 Measures for Comparing Estimated Maps against Ground-truth Data

The BMP is an evaluation measure widely used, not only in the Middlebury's evaluation methodology, but also in other methodologies. It can be seen as binary function using a threshold. A value of 1 pixel is commonly used. This threshold value can be interpreted as associated to an error definition, but it also has historical reasons (i.e. backward compatibility) (Hsieh et al., 1992). Nevertheless, a score of zero in the evaluation produced by the BMP does not necessarily imply that a disparity map is free of errors (Cabezas et al., 2011). The BMP is, in essence, a measure of the quantity of errors in a disparity map, regardless their magnitude. Moreover, the BMP, as well as

other measures used for comparing estimated maps against ground-truth data (i.e. MAD, MSE, PSNR), does not consider the inverse relation between depth and disparity. This is not a trivial issue at all, since even with accurate disparity estimations, location error in a canonical stereo system increases quadratically with depth (Gallup et al., 2008). This fact is illustrated in Figure 4-4, which plots the variation on accuracy estimation according to depth for a Bumblebee® 2-PointGrey stereo camera system (PtGrey, 2012).

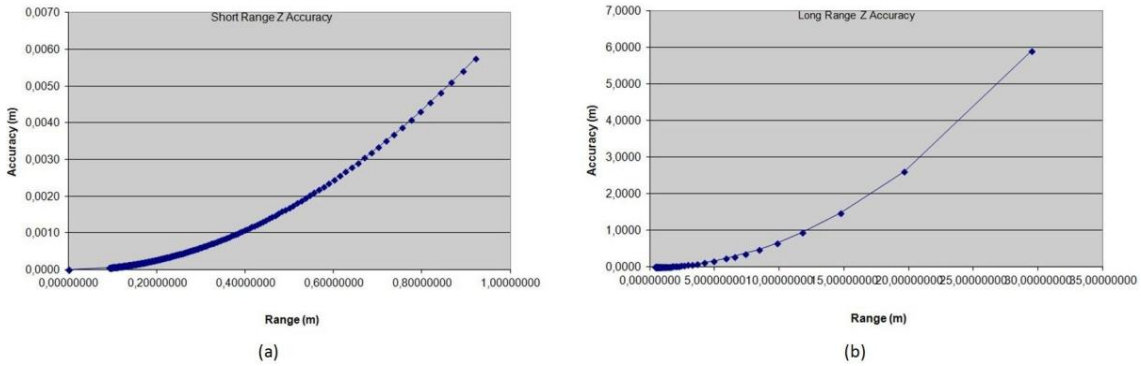


Figure 4-4 Variation on location accuracy estimation according to depth on a commercial stereo camera system (PtGrey, 2012). (a) short range accuracy and (b) long range accuracy.

Thus, in a canonical stereo rig, a disparity estimation error in a distant point may have a larger impact on the final 3D reconstruction, than a disparity estimation error of the same magnitude in a closer point to the stereo camera system (Cabezas et al., 2012c). These issues have to be considered during an evaluation process in several application domains where the final user of a 3D reconstruction is not a human user, such as in smart vehicles or robotic navigation, among others. The relevance of considering both: the inverse relation between depth and disparity, as well as the estimation error magnitude is illustrated in Figure 4-5. On the one hand, Figure 4-5 (a) and Figure 4-5 (c) illustrate how estimation errors of a same magnitude – represented by points p'_r and q'_r respectively – may cause different triangulation errors –represented by points P' and Q', respectively. On the other hand, Figure 4-5 (b) and Figure 4-5 (d) illustrate how a larger estimation error magnitude increases triangulation errors. Consequently, the BMP measure may conceal estimation errors of a large magnitude, and at the same time, it may penalise errors of low impact on a final 3D reconstruction.

Thus, in practice, different disparity maps, with quite similar percentages of BMP, may produce 3D reconstructions of largely different qualities.

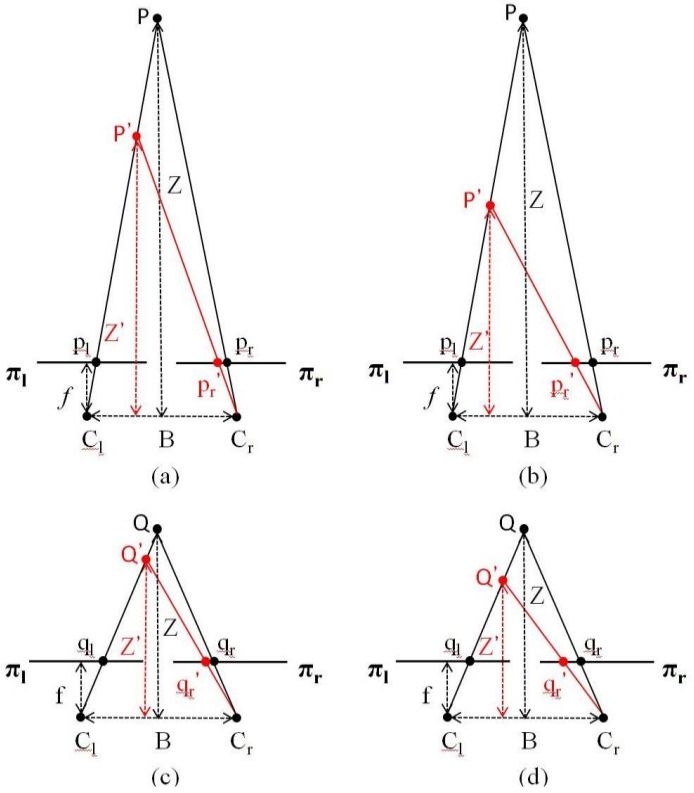


Figure 4-5 Illustration of the Relation between disparity estimation errors and triangulation errors: (a) a small estimation error of a farther point, (b) a large estimation error of a farther point, (c) a small estimation error of a close point, and (d) a large estimation error of a close point.

4.2.4 Evaluation Model and Interpretation of Results

The evaluation model introduced in (Kostlivá et al., 2007) has a solid theoretical support. However, the evaluation of more than one stereo image pair is difficult due to the requirements of weights according to scene's relevance. These weights are used to compute a unique real value by a linear combination of evaluated factors. In practice, assigning such weights may become a really hard and conflictive task (i.e. subjective and difficult to set if a consensus is requiered). Moreover, summarising evaluation results into a single value may hide interesting facts about algorithms performance.

The Middlebury's evaluation methodology is illustrated graphically in Figure 4-5 using a few selected evaluation elements and methods, by following the steps identified in Figure 4-1. In the evaluation model of this methodology, a rank is assigned to each algorithm under evaluation, according to error scores and error criteria. A final ranking is computed by averaging previously established ranks. In this way, the evaluation model of Middlebury's methodology relates ranks to weights. These weights are linearly operated among them in order to produce a single value. In this model, it is assumed that an algorithm with a smaller weight is more accurate than an algorithm with a larger weight. The Middlebury's evaluation model can be modelled as follows.

$$h: (E_{eval}) \rightarrow Ranks, \tag{4.12}$$

where $Ranks$ is a set of real values, $Ranks = \{r \in Ranks \mid r \in \mathbb{R}\}$.

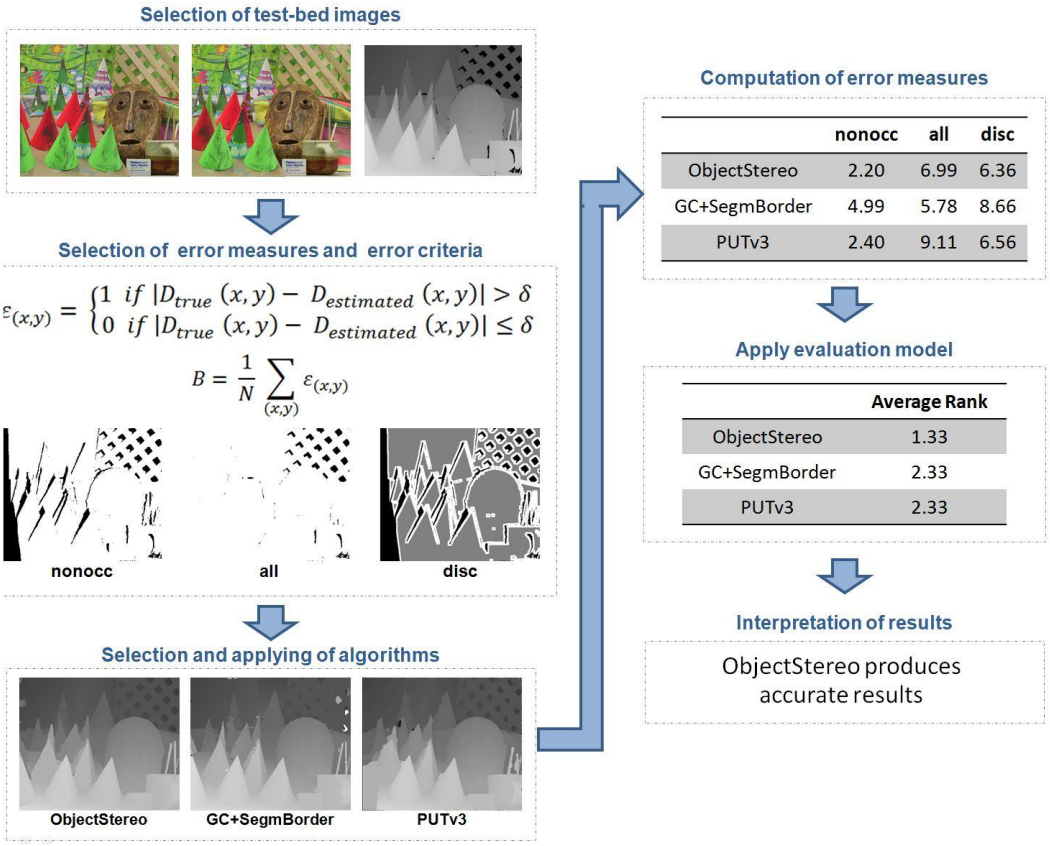


Figure 4-6 Illustration of the steps followed in the Middlebury's evaluation methodology.

However, two or more algorithms may have the same error score under an error criterion. In this case, the rank assigned to these algorithms became arbitrary. This fact

may impact on the final ranking. Additionally, different algorithms may have the same averaged ranking. Nevertheless, it does not mean that these algorithms perform similarly on imagery test-bed. In general terms, the evaluation model does not consider how the evaluation results of different stereo methods can be compared, besides their obvious differences in rankings. Moreover, although it is possible to determine a set of top ranking algorithms based on the Middlebury's methodology, the cardinality of this set is a free parameter. This fact may lead to discrepancies or controversy among researchers about the state-of-the-art on the field. In other word, the Middlebury's may have some ambiguities regarding how evaluation results are computed, and lacking of a well defined way for interpreting results.

4.2.5 Lack of Flexibility

In existing methodologies, evaluation elements and methods to be used in each step of the methodology have been already selected. Thus, evaluation elements and methods are fixed beforehand, and consequently, provided evaluation scenarios are also fixed. Nevertheless, evaluation requirements of a user may vary according to the stage of his/her particular research and development process, or according to the particular algorithmic module under evaluation. For instance, if the user is evaluating an adaptive real-time local stereo correspondence method the evaluation should focus in similar methods, since the comparison against global and off-line methods may be not fair, neither of the interest of the user. In fact, the inclusion of more elements or methods than those required in the evaluation process not only impacts on evaluation results, but also makes difficult the analysis and interpretation of obtained results. Thus, if the evaluation scenario is fixed beforehand it may be not suited to cover different and/or evolving user's requirements.

4.3 Proposal on Evaluation Elements and Methods

The evaluation methodology presented in the thesis follows the steps illustrated in Figure 4-1 (Cabezas & Trujillo, 2011; Cabezas et al. 2012a). It considers most of the weaknesses identified in the previous section, and incorporates some contributions in this regard (Cabezas et al. 2011; 2012b; 2012c; Cabezas & Trujillo, 2012; 2013).

4.3.1 Evaluation Criteria

An alternative interpretation of evaluation criteria is proposed in the thesis. The aim of the reinterpretation is to provide a formal support for their application during the evaluation process. Following the proposed formalisation the multiple inclusion of a point in more than one criterion is avoided. Moreover, it provides guidelines about the possibilities for combining criteria during the evaluation in order to obtain concise and easily interpreted evaluation results. In addition, the proposed formulation allows the evaluation of a new criterion: the disparity assignment in occluded regions. In contrast to the conventional interpretation of evaluation criteria as overlapping segments, the proposed formulation is based on the concept of set partitions. In this way, the reference image, viewed as a set of points, is divided into disjoint sets, which union is equal to the reference image.

For the sake of convenience, the points in the disparity ground-truth which disparity is unknown, and consequently cannot be evaluated, are left out of the formalisation. However these points can be handled without any trouble as an additional partition.

The interpretation of criteria presented in the thesis is formulated as follows.

Let I_l be the reference image composed by N points.

$$I_l = \{p_1, p_2, \dots, p_N\}. \tag{4.13}$$

Let T be a set partition over I_l, as follows:

$$T = \{T_1, T_2, \dots, T_j\}. \tag{4.14}$$

subject to:

$$T_i \subset \mathcal{P}(I_l). \tag{4.15}$$

$$\forall_i \, T_i = \emptyset. \tag{4.16}$$

$$\forall_{i \neq j} \;\; T_i \cap T_j = \emptyset. \tag{4.17}$$

$$\bigcup_{i=1}^{j} T_i = I_l \tag{4.18}$$

$$1 \leq j < N \ \wedge N \in \ \mathbb{N} \tag{4.19}$$

Thus, by definition, a point only belongs to a single criterion. In this way, gathered errors are unequivocally associated to a specific image phenomenon. This formulation imposes a new way to use criteria. Four evaluation scenarios, using the proposed evaluation criteria formulation, are presented below.

4.3.1.1 Evaluation using a Single Criterion

The first case discussed is when there is just one partition, which by definition should be equal to the reference image. It is formulated as follows:

$$T = \{T_1\}. \tag{4.20}$$

This case is termed as the criterion *scene*, and in contrast to the conventional approach, it should be used in isolation. It is associated disparity estimation errors in the entire image. In practice, an evaluation using the criterion scene is an evaluation in the absence of criteria, since it is not possible to associate obtained errors to any image phenomenon. The criterion *scene* is equivalent to the *all* criterion in the Middlebury's methodology.The denomination of *all* will be used in the thesis.

4.3.1.2 Evaluation using *interior, boundary* and *occluded* Criteria

The second evaluation scenario is related to the use of three specific criteria, which are of general interest: *interior*, *boundary* and *occluded* areas. The use of these three criteria allows a comprehensive evaluation of the behaviour of the stereo correspondence method. They are formulated as follows.

$$T = \{T_1 , T_2 , T_3\} \ \wedge \ i = 3. \tag{4.21}$$

$$T_1 = \{p_k \,|p_k \ \in \ I_l \ \wedge \ p_k \ is \ an \ interior \ point\}. \tag{4.22}$$

$$T_2 = \{p_k \,|p_k \ \in \ I_l \ \wedge \ p_k \ is \ a \ boundary \ point\}. \tag{4.23}$$

$$T_3 = \{p_k \,|p_k \in I_l \wedge p_k \, is \, an \, occluded \, point\}\,. \tag{4.24}$$

The *boundary* criterion is associated to those points near to depth discontinuities and occluded regions. Inaccuracies in estimated disparities values of these points may cause artefacts on objects boundaries, producing, for instance, visual discomfort on rendered views. The *interior* criterion considers stereo visible points on smooth surfaces. The criterion *occluded* is associated to reference image points lacking of a corresponding point in the target image. The motivation for evaluating this particular criterion is twofold: firstly, disparity of occluded points cannot be directly estimated from image data since they lack of correspondence. Thus disparities of occluded points should be inferred from disparities of stereo visible points. Secondly, there are several application domains requiring dense maps, independently if captured points are occluded or not. An evaluation on this criterion was not possible in the conventionally used evaluation criteria. These criteria are illustrated for the Teddy reference image in Figure 4-7, and the relation among them is illustrated in Figure 4-8.

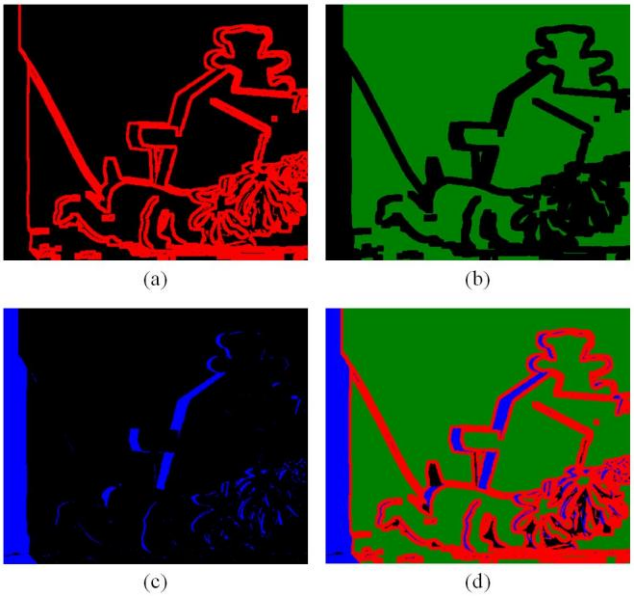


Figure 4-7 Illustration of the evaluation interior, boundary, and occluded criteria using the Teddy left view.

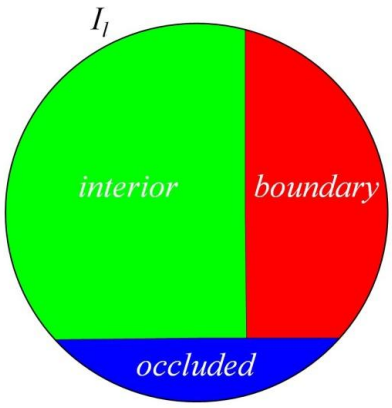


Figure 4-8 Illustration of the relation among the interior, boundary, and occluded criteria, as set partition.

4.3.1.3 Evaluation According to Depth

The third evaluation scenario considers an evaluation based on the depth of the scene using three regions associated to depth ranges: *near*, *mid*, and *far*. The aim of this evaluation is to determine if stereo correspondence methods present some behaviour or biasing for estimating correspondences regarding the depth of the scene (i.e. highly accurate at near disparities, and more susceptible to errors at far disparities, or in contrary, with a homogeneous behaviour regardless the distance of analysed points, among others). In this evaluation scenario, the partitions were computed using a K-means clustering algorithm (Trujillo & Izquierdo, 2005), applied to the disparity maps, with a k input value suited to obtain the specific quantity of clusters wanted. This evaluation scenario is formulated as follows.

$$T = \{T_1, T_2, T_3\} \ \wedge \ i = 3. \tag{4.25}$$

$$T_1 = \{p_k \,|p_k \ \in \ I_l \ \wedge \ p_k \ is \ a \ near \ point\}. \tag{4.26}$$

$$T_2 = \{p_k \,|p_k \ \in \ I_l \ \wedge \ p_k \ is \ a \ mid \ point\}. \tag{4.27}$$

$$T_3 = \{p_k \,|p_k \ \in \ I_l \ \wedge \ p_k \ is \ an \ far \ point\}. \tag{4.28}$$

The depth related criteria masks obtained by clustering the Teddy and the Cones disparity map are illustrated in Figure 4-9, and Figure 4-10, respectively.

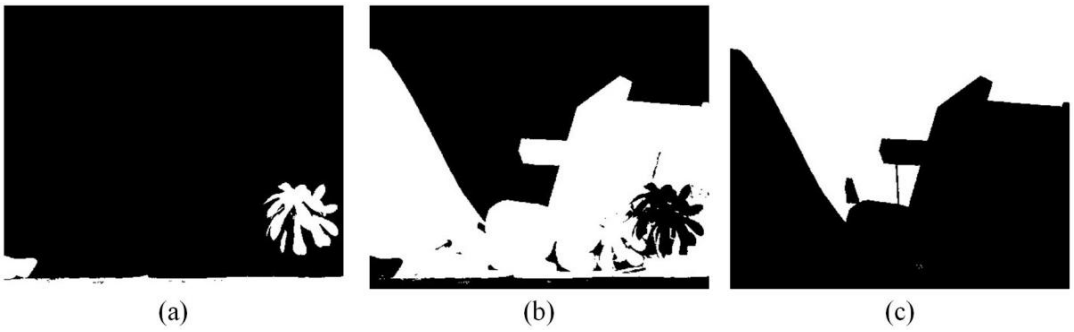


Figure 4-9Depth related evaluation criteria of the Teddy stereo image: (a) near, (b) mid, and (c) far.

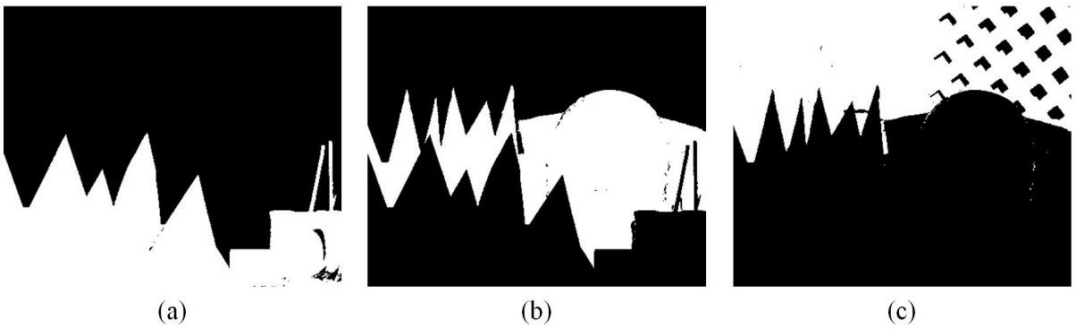


Figure 4-10 Depth related evaluation criteria of the Cones stereo image: (a) near, (b) mid, and (c) far.

4.3.1.4 Evaluation using User Defined Criteria

A fourth evaluation case is related to user defined criteria, according to their interest and/or to some application domain requirements. For instance, some partitions can be related to pedestrians vs. background, or obstacles vs. road, in application such as surveillance or autonomous vehicles, among others (Mark & Gavrila 2006; Gerónimo et al., 2010). Although the number of sets defining a partition may vary, a general case of two sets is addressed here, where, in fact, the interest of the user may be asymmetric. Thus, without loss of generalisation, two regions are considered: region of interest (*roi*) and somewhere else (*elsewhere*). This case is formulated as follows.

$$T = \{T_1 , T_2\}. \tag{4.29}$$

$$T_1 = \{p_k \,|p_k \in I_l \wedge p_k \, is \, a \, roi \, point\}. \tag{4.30}$$

$$T_2 = \{p_k \,|p_k \in I_l \wedge p_k \, is \, an \, elsewhere \, point\}\,. \tag{4.31}$$

Finally, it may be convenient to clarify that, when the scores obtained in any combination of criteria are operated by an evaluation model, the evaluation model it is in fact working over the trade-off achieved by the stereo correspondence method in these three criteria in conjunction.

4.3.2 Comparing Estimated Maps against Ground-truth Data

Among the different components of an evaluation methodology, the set of evaluation measures is a fundamental one. If the selected evaluation measures do not compare properly estimated disparity maps against ground-truth data, then the evaluation process results will be biased. Thus, it is required that evaluation measures properly assert disparity maps taking into account the specific nature of the problem. In this regard, the thesis contains a proposal for two new evaluation measures as well a characterisation of the properties than a measure should fulfil.

4.3.2.1 The Sigma-Z-Error Measure

An error measure termed Sigma-Z-Error (SZE) is proposed in (Cabezas et al., 2011). It is based on the inverse relation between depth and disparity using the error magnitude. In this sense, it aims to measure the final impact of a disparity estimation error, which depends on the true distance between the stereo camera system and the captured point, and on the disparity error magnitude. The SZE is described as follows.

On the one hand, the distance between a point in the captured scene and the camera system can be computed, without loss of generality, based on the information of the stereo rig and the estimated disparity as it is formulated below.

$$Z_{true} = \frac{f * B}{d_{true}}\,, \tag{4.32}$$

where f is the focal length in pixels, B is the baseline in meters, d_{true} is the true disparity value in pixels, and $Z_{true} \in \mathbb{R}$ is the distance along the camera axis in meters.

On the other hand, in practice, an inaccurate Z distance is generated due to a disparity estimation error, as is shown in Equation 4.33.

$$Z_{false} = \frac{f * B}{d_{false}}\ , \tag{4.33}$$

where $Z_{false} \in \mathbb{R}$ is the inaccurate distance estimation and d_{false} is the falsely estimated disparity. Thus, the SZE measure consist in summing the difference between Z_{true}, and Z_{false}, over the entire disparity map (or according to some specific evaluation criterion) based on the information provided by ground-truth data. It is formulated as follows:

$$SZE = \sum_{(x,y)}^{N} \left| \frac{f * B}{D_{true}(x,y) + \mu} - \frac{f * B}{D_{false}(x,y) + \mu} \right|, \tag{4.34}$$

where μ is a small constant for avoiding instability due to missing disparity estimations. Since the SZE measure can be rewritten as the Minkowski distance, it can be seen that it fulfils the properties of a metric. Nevertheless, it is unbounded. In case that the parameters $f$and B are known, they can be incorporated into the measure computation. Otherwise, the computed score is up to a scalar factor.

The proposed measure has three main differences in relation to conventional measures used for comparing disparity maps against ground-truth data:

- The SZE considers the error magnitude as well as the inverse relation between depth and disparity.
- The SZE is related to an error value with a physical interpretation and meaning, which is inherent to the 3D information recovery process.
- The SZE does not require any threshold values.

4.3.2.2 The Bad-Matched-Pixel Relative Error Measure

An error measure termed Bad-Matched-Pixel Relative Error (BMPRE) is proposed in (Cabezas et al., 2012c). The BMPRE is based on two existing measures: the BMP and the MRE. It was designed aiming to incorporate the strengths of these two measures. From the BMP it incorporates the use of a threshold as a way to allow a user to define what he/she considers as mismatch, according to some specific domain. From

the MRE it incorporates the quantification of the error according to the magnitude of the disparity estimation error. Moreover, it also incorporates a variation on the quantification according to the real observation value, which in this context is associated to the inverse relation between depth and disparity.

The BMPRE is defined based on the components formulated below. Let Δ be the magnitude of an estimation error, computed as the absolute difference between the estimated disparity and the true disparity value.

$$\Delta(x,y) = |D_{true}(x,y) - D_{estimated}(x,y)| \,. \tag{4.35}$$

Let ρ be the ratio between Δ and the true disparity value –the relative error according to the ground truth disparity value.

$$\rho(x,y) = \frac{\Delta(x,y)}{D_{true}(x,y)} \,. \tag{4.36}$$

Let τ be a function for avoiding data instability due to missing estimations, defined as follows.

$$\tau(x,y) = \begin{cases} \rho(x,y) & if\ D_{true}(x,y) > 0 \\ 0 & otherwise \end{cases} \,. \tag{4.37}$$

The BMPRE measure is formulated as:

$$BMPRE = \sum_{(x,y)}^{N} \begin{cases} \tau(x,y) & if\ \Delta(x,y) > \psi \\ 0 & if\ \Delta(x,y) \leq \psi \end{cases} \,. \tag{4.38}$$

where ψ is the error threshold in pixels incorporated by a user according to the application domain.

The score produced by the BMPRE can be viewed by a user as a quantification of the global error due to those points exceeding the tolerance threshold. It has two main differences in relation to conventional measures used for comparing disparity maps against ground-truth data:

- The BMPRE is able to consider the error magnitude as well as the inverse relation between depth and disparity, from the properties inherited from the MRE measure.

- The BMPRE offers backward compatibility in relation to already available and published data using the BMP measure. On the one hand, it can be used for evaluating disparity maps in conjunction with the BMP measure if a common threshold value is used. On the other hand, it can be used in isolation in order to quantify the same errors detected by the BMP measure.

In fact, the BMPRE is a measure simple in design (i.e. which makes easier their comprehension by all users) and capable of achieves a proper measurement of error (i.e. by considering both the error magnitude and the inverse relation between depth and disparity). Nevertheless, in the same way that the BMPRE incorporates some of the strengths from the measures on which is based, it also inherits their weaknesses. In this regard, the BMPRE is not metric since it does not fulfil neither the identity (i.e. with any ψ greater than zero), nor the symmetry (i.e. due to the asymmetry in the information content of disparity ground-truth data vs. an estimated disparity map) properties.

The properties of some evaluation measures for comparing estimating disparity maps against disparity ground-thruth are summarised in Table 4-2.

Table 4-2 Properties of evaluation measures for comparing estimated maps against disparity ground-truth data.

Measure	Properties			
	Considers Error Magnitude	*Considers Depth vs. Disparity*	*Advantages*	*Drawbacks*
BMP	No	No	Widely used in related literature Concise error definition	Sensitive to threshold selection, Provides partial information
MAE	Yes	No	Concise interpretation	Not widely used in related literature
MSE	Yes	No	Concise interpretation	Not widely used in related literature
PSNR	Yes	No	Concise interpretation	Not widely used in related literature, Score expresed in dB
MRE	Yes	Yes	Concise interpretation	Sensitive to missing data
SZE	Yes	Yes	Concise interpretation, Theoretical properties	Requires camera system information
BMPRE	Yes	Yes	Concise interpretation Concise error definition	Sensitive to threshold selection

4.3.2.3 A Characterisation of Evaluation Measures

In (Cabezas et al., 2012b) it is shown that the use of different evaluation measures in the evaluation process may lead to contradictory scores of accuracy. This is illustrated in Table 4-2, using different evaluation measures, applied to the results

obtained by selected stereo correspondence methods: ADCensus (Mei et al., 2011) and RDP (Sun et al., 2011), from the Middlebury's benchmark repository (Scharstein & Szeliski, 2012). It can be observed that the BMP and the MRE scores indicate a superior accuracy of the algorithm ADCensus. In contrast, the scores of SZE indicate the inverse relation between these algorithms. This tendency is confirmed by the MSE measure, apart from the *all* criterion in the Teddy and the Cones images,

Table 4-3 Contradictories Evaluation Scores Obtained by Selected Stereo Correspondence Methods According to Different Evaluation Measures

Measure	Teddy			Cones		
	nonocc	all	disc	nonocc	all	disc
	ADCensus					
BMP	4,099	6,216	10,892	2,421	7,254	6,947
SZE	358,779	443,840	162,180	77,358	162,017	46,793
MSE	6,228	7,909	11,712	1,675	4,265	4,888
MRE	0,017	0,022	0,025	0,013	0,021	0,023
	RDP					
BMP	4,836	9,936	12,570	2,535	7,692	7,383
SZE	204,989	348,190	52,043	69,214	113,656	36,888
MSE	4,253	8,649	7,102	1,658	4,437	4,750
MRE	0,021	0,029	0,029	0,014	0,022	0,024

These contradictory scores will impact on obtained evaluation results: if each measure is used in isolation, contradictory results will be produced by a same evaluation model. If contradictories measures are combined, the evaluation model is challenged to find a trade-off among them. This fact entails a problem to an evaluation methodology user which is responsible for selecting evaluation measures. Taking this into account, a characterisation of evaluation measures is presented in (Cabezas et al., 2012b). The characterisation is aimed to assist a user during the selection of a set of measures to be considered in the evaluation process. It is composed by the following attributes:

- **Automatic**: The error measure should be computed without human intervention (i.e. it can be implemented by a computer program). This is an essential attribute of an error measure in a quantitative approach. An error measure based on thresholds can be considered as automatic if such thresholds can (or have to) be fixed prior to its application.

- **Reliable**: The error measure operates without being influenced by external factors and producing always the same output for a particular fixed input (i.e. operates in a deterministic way).
- **Meaningful**: The error measure is intended for a particular purpose, has a concise interpretation related to the phenomenon being analysed, and does not lead to ambiguous results.
- **Unbiased**: The error measure is capable of accomplishing measurements for which it was conceived, and its use allows performing impartially comparisons.
- **Consistent**: The scores produced by an error measure should be compatible (i.e. in agreement of observations) with the scores produced by any other error measure with a common particular purpose.

The level of analysis required, in order to determine if an error measure fulfils a particular attribute, varies according to the attribute being analysed. For instance, although from a theoretical point of view, being automatic can be associated to the halting problem, in practice it can be determined by executing a procedure after the imposition of a reasonable time bound (i.e. proportional to input size and/or lower than an arbitrarily fixed threshold). Once the attribute of being automatic has been established, the next concern is about being reliable. In this situation, it is reasonable to assume that the measure under analysis is reliable until the opposite can be demonstrated by a counter-example. With regard to be meaningful and unbiased, these have to be taken into account during the design (i.e. the formulation) and the implementation (i.e. the conversion to a computer program) phases of the measure. With regard to the consistency, a first attempt to devise a measure of it is conducted. The achieved measure is based on determining the percentage of agreements in obtained results, by applying the evaluation model, whilst the selection of the error measure used is varied. It assumes that each evaluation measure fulfils the former above criteria. Obtained scores by applying evaluation measures are operated by an evaluation model in order to produce evaluation results. In the conducted experimentation the BMP, the SZE, the MSE and the MRE were the considered evaluation measures, and the Middlebury's and the $A^* - Groups$, the evaluation models used. It was observed that

MRE and the SZE shown the highest consistency according to the each one of the models used, respectively.

4.3.3 Evaluation Model and Interpretation of Results

A non-linear evaluation model is proposed in (Cabezas & Trujillo, 2011). The proposed model is termed as $\boldsymbol{A}^*$. It has two main differences with conventionally used models. Firstly, obtained scores are not averaged neither weighted among them in order to obtain a single value of performance. Secondly, it involves a formalisation regarding the interpretation of evaluation results. In this way, the proposed model avoids the operation of incommensurable values, as well as avoids the subjectivity in the interpretation of (quantitative) evaluation results. The proposed model is based on the Pareto Dominance concept. It assumes that all the evaluated criteria share the same relevance. The model is formulated below assuming an inter-technique comparison scenario. This formulation is mathematically equivalent for an intra-technique comparison, on which each different configuration of the method under analysis is handled as a different method.

Let α^* and α be error value vectors.

$$\alpha^*, \alpha \in E_{eval} \tag{4.39}$$

Let $\prec$ be the symbol that denotes the Pareto Dominance relation.

Let $E^*_{eval} | PA^* \subset E_{eval}$ a Pareto optimal set subject to:

$$E^*_{eval} = \{\alpha^* \in E_{eval} \mid \nexists\, \alpha \in E_{eval} \colon \alpha \prec \alpha^*\}. \tag{4.40}$$

Let A^* be a subset of A, $A^* \subseteq A$, subject to:

$$A^* = \left\{a \in A \mid E_{eval(a)} \in E^*_{eval}\right\}. \tag{4.41}$$

Thus, the A^* evaluation model can be formulated as follows.

$$u : (I_{test-bed} \times D_{true} \times R_{criteria} \times A) \rightarrow A^*. \tag{4.42}$$

In brief, the $\boldsymbol{A}^*$ evaluation model aims to find those stereo correspondence methods which associated evaluation errors vectors compose the Pareto set. The model

is termed in relation to the set it aims to compute. Thus, for each algorithm belonging to $A \setminus A^*$ it is possible to find at least one algorithm in the A^* set showing a superior performance according to the Pareto Dominance relation. In addition, the $\boldsymbol{A}^*$ evaluation model considers an interpretation of results which is based on the cardinality of the A^* set, which, by definition, cannot be an empty set. It considers two general cases, which are formulated and described as follows.

$$\begin{cases} if\ |A^*| = 1: single\ method\ of\ superior\ performance \\ otherwise: methods\ of\ comparable\ performance \end{cases}. \quad (4.43)$$

- A Single method of superior performance: if the cardinality of set A^* is equal to 1, it implies that there exists a single stereo method of superior performance over the set of selected methods for evaluation, under the specific evaluation scenario considered.
- A set of stereo methods of comparable performance: if the cardinality of the set A^* is greater than 1, it implies that there exist a set of methods of comparable performance (i.e. not better, neither worst) among them, since their associated vector measure values are incomparable among them.

It is convenient to highlight that obtained evaluation results, for a particular selection of evaluation elements and methods, cannot be not extrapolated for a different evaluation scenario (i.e. to imagery test-bed captured in different conditions, to other possible criteria, or to comparisons including different stereo methods, among others).

In (Cabezas et al., 2012a) it is pointed that the $\boldsymbol{A}^*$ evaluation model does not consider an evaluation scenario on which a user is interested in an exhaustive evaluation of the entire set stereo correspondence methods, instead of only determining about a single method or a group of methods of superior performance overall. An extension to the $\boldsymbol{A}^*$ model, which takes into account the above consideration, is proposed in (Cabezas et al., 2012a). The extension is based on iteratively evaluating the entire set of stereo correspondence methods by computing groups of comparable accuracy. Thus, each obtained group is associated to a different level of accuracy. The composition of each group is unambiguously determined based on the Pareto Dominance relation. The extension of the evaluation model is termed as $\boldsymbol{A}^* - Groups$. It can be defined as follows.

Let A_l^+ be a non-empty set of stereo correspondence methods, subset of A, subject to:

$$A_l^+ \neq \{\} \wedge A_l^+ \cap A_l^* \wedge A_l^+ \cup A_l^* \subseteq A_{l-1}^+. \quad (4.43)$$

A discrete label l, associated to the respective group is computed by the extended model.

$$u' : (I_{test-bed} \times D_{true} \times R_{criteria} \times A_l^+) \rightarrow A_l^* \cup A_{l+1}^+ . \quad (4.44)$$

subject to:

$$A_l^* = \{a \in A^+ \mid E_{eval(a)} = E_{eval_l}^*\} \quad (4.45)$$

$$\forall \alpha \in E_{eval_{l+1}}^* \exists \alpha^* \in E_{eval_l}^* \mid \alpha^* \prec \alpha. \quad (4.46)$$

In brief, the $\boldsymbol{A^* - Groups}$ model updates the set of stereo methods under evaluation by iteratively subtracting the stereo method or methods of comparable performance among them, and superior to the rest of methods, and conducting again the evaluation until reaching an empty set. The criteria applied for the interpretation of results from the evaluation model still hold, but applied in this case to the A_1^* group.

4.3.3.1 A Method for Reducing the Cardinality Pareto Front

The $\boldsymbol{A^* - Groups}$ evaluation model computes a Pareto front based on evaluation score vectors. Thus, a user should analysing obtained results in order to select a single stereo method to be used (or perhaps implemented). However, the cardinality of the Pareto front, as well as the multidimensionality obtained results may overload judging capabilities of users, which, consequently, requires assistance in his/her decision making process. A method for reducing the cardinality of the Pareto front is proposed in (Cabezas & Trujillo, 2012). In the proposed method, the selection of a solution from the Pareto front is addressed as a MOP, based on two utility functions and the Pareto dominance relation. The utility functions are adapted from (Bentley & Wakefield, 1997) in order to avoid the use of weights. These functions are computed over the vectors composing the Pareto front from which a solution should be selected. They are

computed over the vectors composing the Pareto front from which a solution should be selected.

Thus, the proposed method consists in finding the vector $s = (f_1(x), f_2(x), \dots, f_K(x))^T$ that optimises the following equation:

$$Min_s\, u(s) = \left(u_1(s), u_2(s)\right)^T, \tag{4.47}$$

subject to:

$$\mathrm{s} \in \mathrm{PF}^* , \tag{4.48}$$

where $\mathrm{u_l}: \mathbb{R}^{\mathrm{K}} \to \mathbb{R}\ (\mathrm{l} = 1, 2)$ are the objective functions, and PF^* is the Pareto front.

Let u_1 be the sum of ranks assigned to $\mathrm{f_k}: \mathbb{R}^{\mathrm{n}} \to \mathbb{R}\ (\mathrm{k} = 1, \dots, \mathrm{K})$ in the Pareto front:

$$u_1\,(s) = \sum_{k=1}^{K} Rank(f_k(x)). \tag{4.49}$$

Let u_2 be the sum of ratios of $\mathrm{f_k}: \mathbb{R}^{\mathrm{n}} \to \mathbb{R}\ (\mathrm{k} = 1, \dots, \mathrm{K})$ in the Pareto front:

$$u_2\,(s) = \sum_{k=1}^{K} \frac{(f_k(x) - Min(f_k(x)))}{(Max(f_k(x) - Min(f_k(x)))}, \tag{4.50}$$

where $\mathrm{Min}(\mathrm{f_k}(\mathrm{x}))$ and $\mathrm{Max}(\mathrm{f_k}(\mathrm{x}))$, are the minimum and the maximum score of the kth objective, respectively.

The lowest sum of ranks is associated with the solution that, comparatively with other solutions in the Pareto front, minimises most of involved objectives, whilst the lowest sum of ratios is associated with the solution with the best objective function values. The selection of a final solution may be based on the above criteria, which are problem context independent. Thus, the set of possible solutions to select from is turned into a set of a small cardinality, or even into a single solution, depending on data, by the proposed method. The set that corresponds to a reduction of the original PF^* set is denoted as RPF^*. In addition, the reduction of cardinality allows the use of a parallel coordinates plotting diagram (Brockhoff et al., 2006) as a visualisation tool such as for

assisting a solution selection. Moreover, the values computed by the u_2 function can be used for plotting the diagram.

4.4 Chapter Summary

- The formulation of evaluation criteria as sets partition allows a gathering of evaluation scores on which an estimation error is considered only once. In this way, reported scores can be unbiasedly associated to a. single image phenomenum. Moreover, the presented formulation allows an evaluation on interesting areas for diverse application domains, such as occluded areas.
- The BMP is a measure of the quantity of errors in a disparity map. It does not consider the magnitude of an error, neither the inverse relation between depth and disparity. Consequently, its use, as the exclusive measure for evaluating disparity maps, may not provide enough information on theaccurcy of estimated disparity maps.
- The BMPRE measure was designed in order to exploit the error definition concept from the BMP, and the quantification of disparity estimation error magnitude and inverse relation between depth and disparity from the MRE. It can be used in conjunction to the BMP for a better analysis and understanding of already available evaluation results data, as well as in insolation, in order to properly evaluate the impact of estimaton errors on depth calculations.
- The SZE distance is a metric inherently related to the 3D information recovery process. This condition offers theoretical advantages over other measures used in disparity maps evaluation. This measure assumes that used disparity ground-truth data is highly reliable, and.is suited to be used during an evaluation process requiring high precision.
- A multi-objective optimisation problem involves a decision-making process in a multi-dimensional space. Once a Pareto front (or an approximation to it) has been computed, selecting a solution from it may overload the judging capabilities of a decision maker. The proposed method for reducing the cardinality of the obtained Pareto front addressed the decision making as a multi-objective

optimisation problem, based on two utility functions. It is motivated in the context of a methodology for evaluating disparity maps, but it can be used in any other multi-objective optimisation problem.

CHAPTER 5.

EXPERIMENTAL EVALUATION

The proposed evaluations elements and methods of the thesis are motivated in this chapter, and used, as the discussion progresses, in order to exemplify their relevance on assessing the impact of mismatches on depth calculations. They are compared against the well-known Middlebury's evaluation methodology (Scharstein & Szeliski, 2002; 2003; 2012). The selection of stereo methods to be compared is conducted taking into account common characteristics of them.

Chapter Contents

5.1 An Adaptive and Interactive Evaluation Framework

An on-line evaluation framework was developed in order to validate the presented proposals. It is available at htpp://ivancabezas.com/stereo_eval/. It is adaptive and interactive in the sense that a user may configure an evaluation scenario, by selecting among different evaluation elements and methods. A user configured evaluation process, using the developed environment, is illustrated in Figure 5-1 to Figure 5-6, following the steps identified in Figure 4-1. The imagery test-bed is selected by a user among the Middlebury's benchmark stereo images (i.e. all of them or a subset). Figure 5-1 illustrates the selection of imagery test-bed.

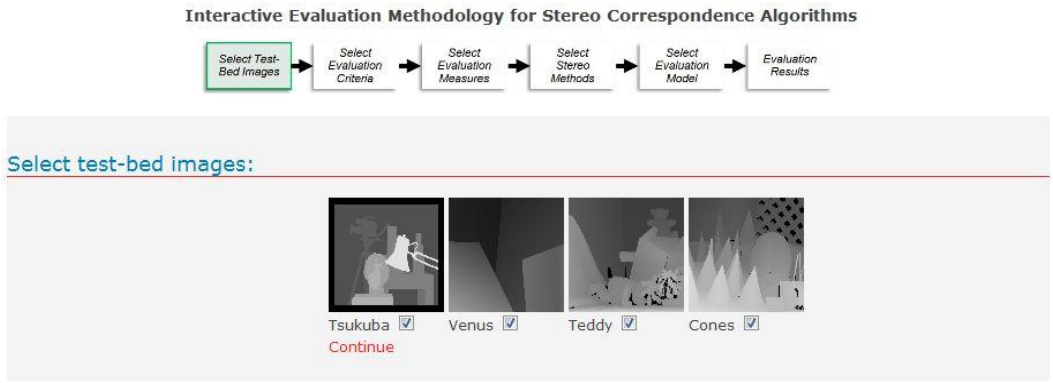


Figure 5-1 Screen shot of the on-line evaluation framework: selection of imagery test-bed.

With regard to evaluation criteria, the selection may include not only the proposed criteria, but also those conventionally used. Figure 5-2 illustrates the selection of evaluation criteria.

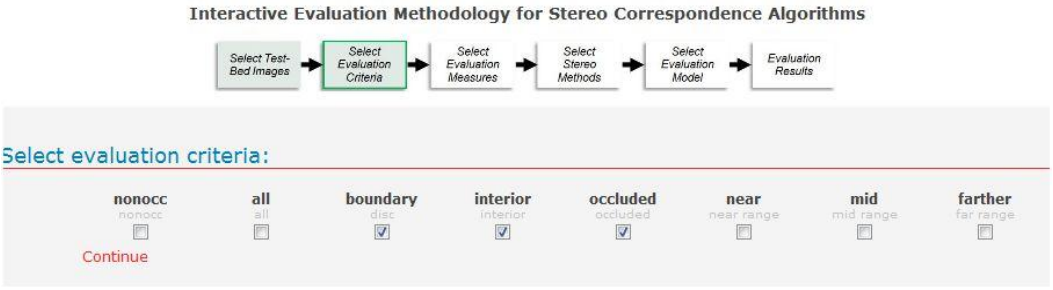


Figure 5-2 Screen shot of the on-line evaluation framework: selection of evaluation criteria.

The selection of evaluation measures is performed among the SZE, the BMPRE, the BMP, the MAE, the MSE, and the MRE. It is illustrated in Figure 5-3.

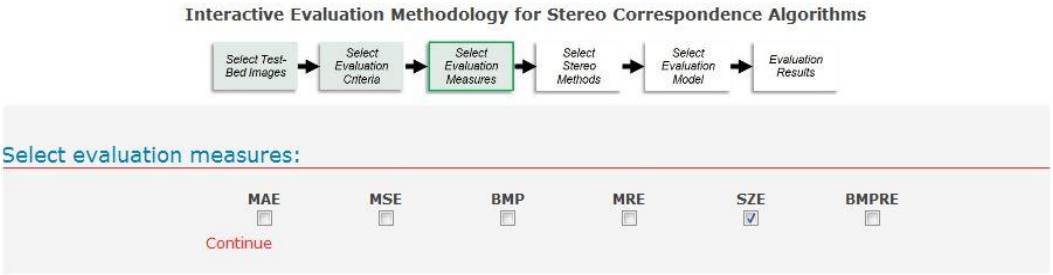


Figure 5-3 Screen shot of the on-line evaluation framework: selection of evaluation measures.

The stereo methods reported to the Middlebury's repository (Scharstein & Szeliski, 2012), can be selected, entirely or partially, to be compared. In this case,

evaluation scores are already stored in a database, so an execution of methods is not required. Figure 5-4 illustrates the selection of stereo methods.

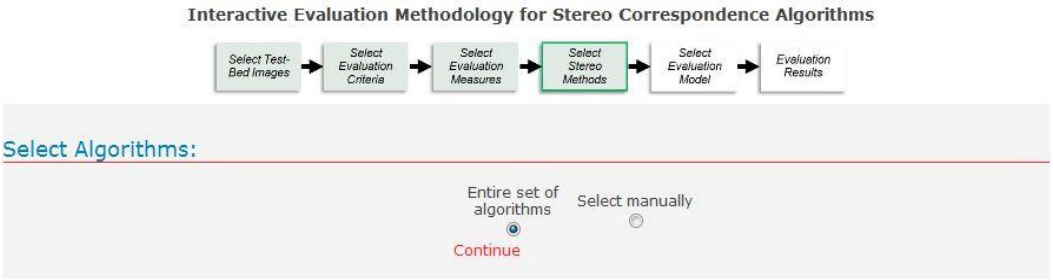


Figure 5-4 Screen shot of the on-line evaluation framework: selection of stereo methods.

Two evaluation models can be used: the $A^* - Groups$ model or the Middlebury's model. Figure 5-5 illustrates the selection of an evaluation model for comparing evaluation scores, according to already selected elements and methods.

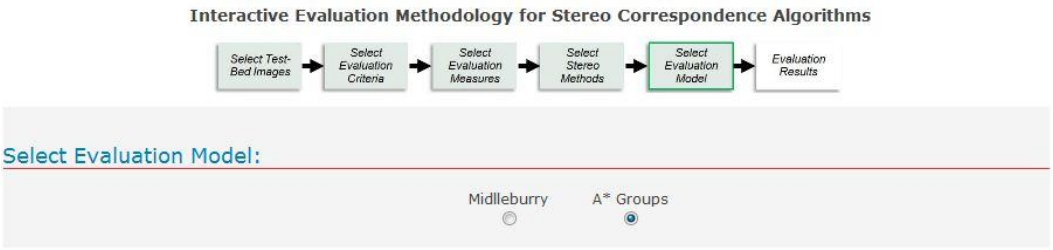
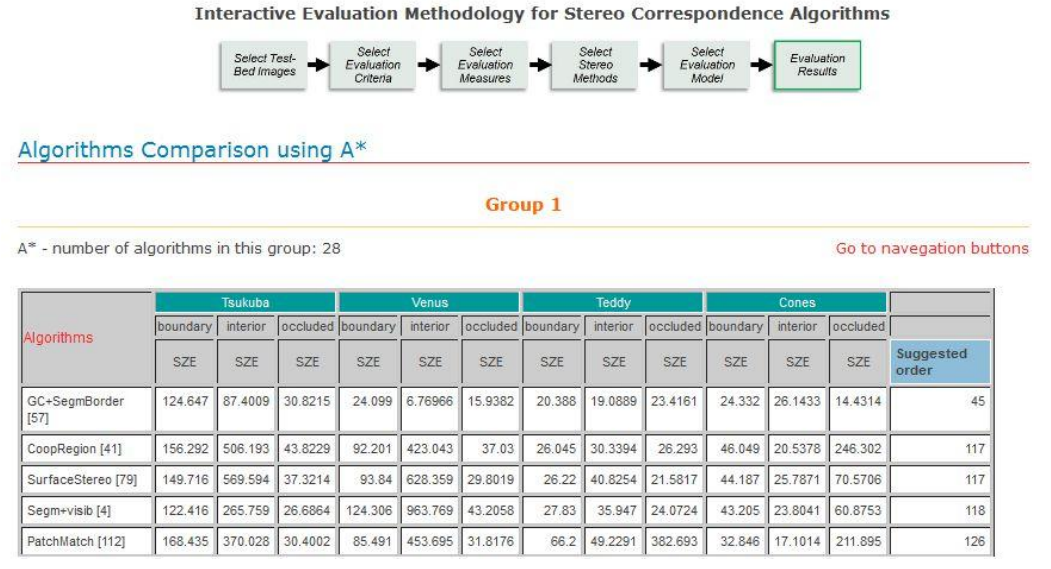


Figure 5-5 Screen shot of the on-line evaluation framework: selection of the evaluation model.



Algorithms	Tsukuba			Venus			Teddy			Cones			
	boundary	interior	occluded	boundary	interior	occluded	boundary	interior	occluded	boundary	interior	occluded	
	SZE	SZE	SZE	SZE	SZE	SZE	SZE	SZE	SZE	SZE	SZE	SZE	Suggested order
GC+SegmBorder [57]	124.647	87.4009	30.8215	24.099	6.76966	15.9382	20.388	19.0889	23.4161	24.332	26.1433	14.4314	45
CoopRegion [41]	156.292	506.193	43.8229	92.201	423.043	37.03	26.045	30.3394	26.293	46.049	20.5378	246.302	117
SurfaceStereo [79]	149.716	569.594	37.3214	93.84	628.359	29.8019	26.22	40.8254	21.5817	44.187	25.7871	70.5706	117
Segm+visib [4]	122.416	265.759	26.6864	124.306	963.769	43.2058	27.83	35.947	24.0724	43.205	23.8041	60.8753	118
PatchMatch [112]	168.435	370.028	30.4002	85.491	453.695	31.8176	66.2	49.2291	382.693	32.846	17.1014	211.895	126

Figure 5-6 – Screen shot of the on-line evaluation framework: obtained evaluation results.

An illustration of obtained results under user's selection is shown in Figure 5-6.

5.2 Near Real-Time, Real-Time, and GPU Based Stereo Methods Comparison

In this evaluation scenario, the goal is comparing stereo methods with near real-time or real-time performance, and/or GPU-based. The selected methods are listed in Table 5-1.

Table 5-1 Selected stereo methods of near real-time and real-time performance

Method	Type	Reference
ADCensus	Global, GPU-based, Near Real-Time	(Mei et al., 2011)
CostFilter	Local, GPU-based, Near Real-Time	(Rhemann et al., 2011)
DCBGrid	Global, GPU-based, Real-Time	(Richardt et al., 2010)
FastAggreg	Local, Near Real-Time	(Tombari et al., 2008)
GeoDif	Local, GPU-based, Near Real-Time	(De-Maeztu et al., 2012)
OptimizedDP	Global, Near Real-Time	(Salmen et al., 2009)
PlaneFitBP	Global, GPU-based, Near Real-Time	(Yang et al., 2008)
RealTimeABW	Local, Real-Time	(Gupta & Sho, 2010)
RealtimeBFV	Local, GPU-based, Real-Time	(Zhang et al., 2009b)
RealtimeBP	Global, GPU-based, Real-Time	(Yang et al., 2006)
RealTimeGPU	Global, GPU-based, Real-Time	(Wang et al., 2006b)
RealtimeVar	Global, Near Real-Time	(Kosov et al., 2009)
ReliabilityDP	Global, GPU-based, Real-Time	(Gong & Yang, 2005)
RTAdaptWgt	Local, GPU-based, Real-Time	(Kowalczuk et al., 2012)
RTCensus	Local, GPU-based, Real-Time	(Humenberger et al., 2010)
RT-ColorAW	Global, GPU-based, Real-Time	(Chang et al., 2011)
TwoWin	Local, Real-Time	(Gupta & Sho, 2010)

Evaluation results obtained by applying the elements and methods of the Middlebury's methodology (i.e. *nonocc*, *all* and *disc* as evaluation criteria, BMP as evaluation measure with δ equals to 1 pixel, and the Middlebury's evaluation model) to selected methods are shown in Table 5-2. An interpretation of results based on this evaluation model can be stated as follows: the ADCensus method is the top performing method, the PlaneFitMethod method is superior to the CostFilter method, the CostFilter method is superior to the GeoDif method, and so on.

5.2.1 Selection of Evaluation Criteria

According to the proposed formulation of evaluation criteria as sets partition, the guidelines for using and combining evaluation criteria are as follows:

Table 5-2 Evaluation results by Middlebury's methodology stereo methods of near real-time and real-time performance.

Method	Rank	Avg.	Tsukuba			Venus			Teddy			Cones		
			nonocc	*all*	*disc*	*nonocc*	*all*	*disc*	*nonocc*	*all*	*disc*	*nonocc*	*all*	*disc*
			BMP											
ADCensus	1	1,17	1,074	1,485	5,731	0,087	0,254	1,148	4,099	6,216	10,892	2,421	7,254	6,947
PlaneFitBP	2	3,67	0,974	1,827	5,256	0,169	0,506	1,708	6,645	12,144	14,712	4,172	10,688	10,604
CostFilter	3	3,92	1,515	1,847	7,606	0,204	0,391	2,419	6,163	11,796	16,043	2,706	8,245	7,663
GeoDif	4	5	1,883	2,349	7,638	0,378	0,818	3,017	5,993	11,313	13,291	2,84	8,33	8,087
RTAdaptWgt	5	5,92	1,449	1,99	7,587	0,402	0,807	3,378	7,648	13,276	16,24	3,481	9,344	8,813
RT-ColorAW	6	6,92	1,402	3,075	5,814	0,721	1,715	3,795	6,685	13,953	15,31	4,03	11,869	10,219
RealTimeABW	7	8,83	1,264	1,672	6,827	0,328	0,651	3,558	10,71	18,259	23,336	4,811	12,622	10,731
RealtimeBP	8	9,25	1,488	3,396	7,872	0,772	1,899	9,004	8,717	13,242	17,171	4,611	11,641	12,444
RealtimeBFV	9	9,5	1,714	2,22	6,745	0,546	0,868	2,884	9,902	15,002	19,501	6,662	12,338	13,378
FastAggreg	10	10,58	1,16	2,114	6,061	4,029	4,749	6,433	9,041	15,165	20,214	5,37	12,603	11,935
RealtimeVar	11	10,83	3,332	5,478	16,846	1,154	2,346	12,827	6,178	13,096	17,324	4,662	11,695	13,668
RTCensus	12	11,83	5,082	6,245	19,227	1,582	2,415	14,241	7,964	13,815	20,29	4,099	9,536	12,221
OptimizedDP	13	12	1,972	3,782	9,797	3,325	4,742	12,97	6,529	13,889	16,576	5,166	13,737	13,438
ReliabilityDP	14	13,08	1,359	3,386	7,245	2,354	3,484	12,22	9,818	16,867	19,54	12,884	19,945	19,71
RealTimeGPU	15	13,17	2,049	4,223	10,64	1,92	2,982	20,266	7,23	14,38	17,573	6,412	13,7	16,462
TwoWin	16	13,33	2,246	3,085	11,59	0,924	1,31	7,533	10,712	15,8	23,598	8,249	13,455	16,608
DCBGrid	17	14	5,904	7,259	21,045	1,347	1,91	11,186	10,487	17,234	22,173	5,342	11,947	14,876

- A combination of criteria: where the union of considered criteria is, at most, equal to the whole set of points under evaluation.

- A specific criterion in isolation: associated to the aim of a particular criterion, where the *all* criterion is a special case allowing a general (i.e. concise) evaluation, but unrelated to image phenomena.

Taking into account the above guidelines, the conventional use of the Middlebury's evaluation criteria (i.e. the *all* criterion used with any other criterion, as well as the *nonocc* criterion used with the *disc* criterion) violates the proposed principles for evaluation criteria, since they are not disjoint among them. The problematic generated by this issue is illustrated on Table 5-3, Table 5-4, Table 5-5 and Table 5-6. In these tables, the quantity of estimation errors using the BMP measure, under conventional evaluation criteria, and under the proposed criteria, are shown in the columns entitled Middlebury's Evaluation Criteria, and Proposed Evaluation Criteria, respectively. It can be observed that, in Table 5-3 to Table 5-6, the quantity of badly matched pixels of the *nonocc* criterion is always larger than the quantity associated to the *disc* criterion. Intuitively, this can be interpreted as that image phenomena associated to the *nonocc* criterion is more challenging than image phenomena associated to the *disc* criterion. However, this is not the case.

Table 5-3 Quantity of badly matched pixels for the Tsukuba image estimated by selected methods of near real-time and real-time performance

Method	Tsukuba					
	Middlebury's Evaluation Criteria			Proposed Evaluation Criteria		
	all	*nonocc*	*disc*	*occluded*	*interior*	*boundary*
ADCensus	1302	918	905	385	13	905
PlaneFitBP	1602	832	830	770	2	830
CostFilter	1620	1294	1201	325	93	1201
GeoDif	2060	1609	1206	451	403	1206
RTAdaptWgt	1745	1238	1198	507	40	1198
RT-ColorAW	2697	1198	918	1499	280	918
RealTimeABW	1466	1080	1078	386	2	1078
RealtimeBP	2978	1271	1243	1707	28	1243
RealtimeBFV	1947	1464	1065	482	399	1065
FastAggreg	1854	991	957	863	34	957
RealtimeVar	4804	2847	2660	1957	187	2660
RTCensus	5477	4342	3036	1135	1306	3036
OptimizedDP	3317	1685	1547	1632	138	1547
ReliabilityDP	2969	1161	1144	1808	17	1144
RealTimeGPU	3703	1751	1680	1953	71	1680
TwoWin	2705	1919	1830	786	89	1830
DCBGrid	6366	5044	3323	1322	1721	3323

Table 5-4 Quantity of badly matched pixels for the Venus image estimated by selected methods of near real-time and real-time performance

Method	Veuns					
	Middlebury's Evaluation Criteria			Proposed Evaluation Criteria		
	all	*nonocc*	*disc*	*occluded*	*interior*	*boundary*
ADCensus	382	128	121	253	7	121
PlaneFitBP	760	249	180	511	69	180
CostFilter	588	301	255	287	46	255
GeoDif	1229	558	318	672	240	318
RTAdaptWgt	1213	593	356	620	237	356
RT-ColorAW	2577	1064	400	1514	664	400
RealTimeABW	978	484	375	494	109	375
RealtimeBP	2854	1139	949	1715	190	949
RealtimeBFV	1304	805	304	499	501	304
FastAggreg	7137	5943	678	1194	5265	678
RealtimeVar	3526	1702	1352	1823	350	1352
RTCensus	3629	2334	1501	1296	833	1501
OptimizedDP	7126	4905	1367	2222	3538	1367
ReliabilityDP	5236	3472	1288	1763	2184	1288
RealTimeGPU	4481	2832	2136	1649	696	2136
TwoWin	1969	1363	794	606	569	794
DCBGrid	2870	1987	1179	883	808	1179

Table 5-5 Quantity of badly matched pixels for the Teddy image estimated by selected methods of near real-time and real-time performance

Method	Teddy					
	Middlebury's Evaluation Criteria			Proposed Evaluation Criteria		
	all	*nonocc*	*disc*	*occluded*	*interior*	*boundary*
ADCensus	10278	6052	4413	4226	1639	4413
PlaneFitBP	20079	9811	5961	10268	3851	5961
CostFilter	19504	9100	6500	10404	2600	6500
GeoDif	18705	8849	5385	9857	3464	5385
RTAdaptWgt	21951	11292	6580	10659	4712	6580
RT-ColorAW	23070	9870	6203	13200	3667	6203
RealTimeABW	30190	15813	9455	14377	6358	9455
RealtimeBP	21895	12871	6957	9024	5914	6957
RealtimeBFV	24805	14620	7901	10185	6719	7901
FastAggreg	25074	13349	8190	11725	5159	8190
RealtimeVar	21653	9122	7019	12532	2103	7019
RTCensus	22842	11759	8221	11083	3538	8221
OptimizedDP	22965	9640	6716	13324	2924	6716
ReliabilityDP	27889	14496	7917	13392	6579	7917
RealTimeGPU	23776	10675	7120	13101	3555	7120
TwoWin	26124	15816	9561	10308	6255	9561
DCBGrid	28495	15484	8984	13011	6500	8984

The above misinterpretation is due to the fact that the *disc* criterion is a subset of the *nonocc* and the *all* criteria. In this way, errors under the *disc* criterion are counted, twice, under the *nonocc* criterion, and even a third time, under the *all* criterion. Thus, an error is not counted only once, and some errors are counted more time than others.

Table 5-6 Quantity of badly matched pixels for the Cones image estimated by selected methods of near real-time and real-time performance

Method	Cones					
	Middlebury's Evaluation Criteria			Proposed Evaluation Criteria		
	all	*nonocc*	*disc*	*occluded*	*interior*	*boundary*
ADCensus	11847	3484	3278	8363	206	3278
PlaneFitBP	17456	6005	5004	11451	1001	5004
CostFilter	13466	3895	3616	9571	279	3616
GeoDif	13605	4087	3816	9517	271	3816
RTAdaptWgt	15261	5010	4159	10251	851	4159
RT-ColorAW	19385	5800	4822	13584	978	4822
RealTimeABW	20614	6924	5064	13690	1860	5064
RealtimeBP	19012	6636	5872	12376	764	5872
RealtimeBFV	20151	9588	6313	10562	3275	6313
FastAggreg	20583	7729	5632	12855	2097	5632
RealtimeVar	19100	6710	6450	12391	260	6450
RTCensus	15574	5900	5767	9675	133	5767
OptimizedDP	22435	7435	6341	15000	1094	6341
ReliabilityDP	32574	18543	9301	14031	9242	9301
RealTimeGPU	22375	9229	7768	13146	1460	7768
TwoWin	21975	11872	7837	10102	4035	7837
DCBGrid	19512	7689	7020	11823	669	7020

In contrast, under the proposed criteria, errors are counted just once, allowing a clear interpretation of errors with regard to image phenomena. In fact, in the case of the Tsukuba image, as it is shown in Table 5-3, it can be observed that for several stereo methods (i.e. ADCensus, PlaneFitBP, RealTimeABW, RealTimeBP, ReliabilityDP, among others) the quantity of estimation errors under the *interior* criterion are negligible in comparison to errors under the *boundary* criterion. In the case of the Venus image, as it is shown in Table 5-4, it can be observed that for the ADCensus method the quantity of errors in the *interior* criterion is negligible, whilst for several methods (i.e. such as PlaneFitBP, CostFilter, RealtimeBP, RealtimeVar, among others) the quantity of badly matched pixels under the *interior* criterion is notoriously lower than the quantity of errors associated to the *boundary* criterion. Moreover, it can be observed that, comparatively, for methods such as FastAggreg, OptimizedDP, and ReliabilityDP, image phenomena associated to the *interior* criterion are more challenging than *boundary* criterion. In the case of the Teddy image, as it is shown in Table 5-5, the difference between errors in *boundary* and *interior* are around a third or a half part for several stereo methods (i.e. ADCensus, CostFilter, RealTimeVar, RTCensus, and RealTimeGPU). Moreover, taking as example, just mentioned methods, it can be observed that in the case of the Cones

image, shown in Table 5-6, the proportion of errors under the *interior* criterion is considerably lower (i.e. one order of magnitude of difference in most of cases) than errors under the *boundary* criterion. From data shown in Table 5-3, to Table 5-6 it can be inferred that disparity estimation in areas near to depth boundaries is more challenging for selected stereo methods than disparity estimation on smooth surfaces, disregarding the lack of texture and the presence of repetitive patterns (i.e. at least on used test-bed imagery or on images captured in similar conditions).

On the other hand, the above analysis may reveal another weakness of the BMP measure: the obtained score, being expressed as a percentage, is a relative value associated to an unknown factor, since masks' sizes upon which percentages are computed, are unknown for most users, as well as considerably different for each criterion.

The use of the proposed criteria may have an impact on the assigned ranks by the Middlebury's evaluation model, and consequently in obtained evaluation results. In addition, and more important than possible changes on assigned ranks, it may reveal useful information about the aspect or aspects on which the method requires adjustments and/or improvements. Evaluation of selected methods under the *interior*, the *occluded*, and the *boundary* criteria, using Middlebury's evaluation model is shown in Table 5-7. This selection of multiple criteria corresponds to a case on which evaluation's goal is to analyse the behaviour of selected stereo methods with regard to different image phenomena. The Table 5-8 shows evaluation results by combining *interior* and *boundary* criteria (i.e. a general case assuming that some selected methods may not have an occlusion model). It can be observed by comparing shown results in Table 5-2 against those shown in Table 5-7 and Table 5-8, that:

- With regard to Table 5-7, the use of the proposed criteria produces some slightly different results for most of selected stereo methods (i.e. a difference of one or two positions on ranks for PlaneFitBP, CostFilter, RTCensus, RealTimeGPU, DCBGrid, among others), and a more significant change for a few of them (i.e. a difference of three or more positions on ranks RT-ColorAW, RealTimeBFV, TwoWin, ReliabilityDP), considering the small cardinality of the set of methods being compared.

Table 5-7 Evaluation of selected methods of near real-time and real-time performance under the proposed criteria and using Middlebury's evaluation model

Method	Rank	Avg.	Tsukuba			Venus			Teddy			Cones		
			boundary	*interior*	*occluded*	*boundary*	*interior*	*occluded*	*boundary*	*interior*	*occluded*	*boundary*	*interior*	*occluded*
			BMP											
ADCensus	1	1,42	5,731	0,019	17,006	1,148	0,006	9,101	10,892	1,530	23,885	6,947	0,214	43,119
CostFilter	2	4,33	7,606	0,134	14,438	2,419	0,034	10,365	16,043	2,427	58,803	7,663	0,287	49,348
PlaneFitBP	3	5,00	5,256	0,003	34,101	1,708	0,050	18,490	14,712	3,595	58,029	10,604	1,034	59,046
GeoDif	4	5,67	7,638	0,579	19,973	3,017	0,175	24,269	13,291	3,233	55,706	8,087	0,281	49,064
RTAdaptWgt	5	6,83	7,587	0,057	22,454	3,378	0,173	22,391	16,240	4,399	60,239	8,813	0,880	52,849
RealtimeBFV	6	8,83	6,745	0,573	21,391	2,884	0,366	18,021	19,501	6,273	57,554	13,378	3,387	54,452
RealTimeABW	7	8,92	6,827	0,003	17,095	3,558	0,080	17,840	23,336	5,935	81,258	10,731	1,923	70,590
RealtimeBP	8	9,08	7,872	0,040	75,598	9,004	0,139	61,936	17,171	5,519	51,003	12,444	0,791	63,805
RT-ColorAW	9	9,42	5,814	0,402	66,386	3,795	0,485	54,641	15,310	3,424	74,600	10,219	1,011	70,044
FastAggreg	10	10,25	6,061	0,049	38,220	6,433	3,845	43,084	20,214	4,815	66,269	11,935	2,168	66,275
RTCensus	11	10,50	19,227	1,875	50,266	14,241	0,608	46,804	20,290	3,302	62,641	12,221	0,137	49,884
TwoWin	12	10,83	11,590	0,128	34,810	7,533	0,415	21,849	23,598	5,839	58,260	16,608	4,172	52,080
RealtimeVar	13	10,92	16,846	0,268	86,670	12,827	0,256	65,800	17,324	1,963	70,830	13,668	0,269	63,888
RealTimeGPU	14	12,42	10,640	0,102	86,448	20,266	0,508	59,588	17,573	3,318	74,046	16,462	1,510	67,780
OptimizedDP	15	12,50	9,797	0,198	72,276	12,970	2,583	80,210	16,576	2,729	75,307	13,438	1,131	77,345
DCBGrid	16	12,50	21,045	2,471	58,547	11,186	0,590	31,889	22,173	6,067	73,538	14,876	0,692	60,959
ReliabilityDP	17	13,58	7,245	0,024	80,071	12,220	1,594	63,705	19,540	6,142	75,685	19,710	9,555	72,338

Table 5-8 Evaluation of selected methods of near real-time and real-time performance under interior and boundary criteria, using Middlebury's evaluation model

Method	Rank	Avg.	Tsukuba		Venus		Teddy		Cones	
			boundary	*interior*	*boundary*	*interior*	*boundary*	*interior*	*boundary*	*interior*
			BMP							
ADCensus	1	1,50	5,731	0,019	1,148	0,006	10,892	1,530	6,947	0,214
PlaneFitBP	2	4,38	5,256	0,003	1,708	0,050	14,712	3,595	10,604	1,034
CostFilter	3	4,88	7,606	0,134	2,419	0,034	16,043	2,427	7,663	0,287
GeoDif	4	6,38	7,638	0,579	3,017	0,175	13,291	3,233	8,087	0,281
RTAdaptWgt	5	6,88	7,587	0,057	3,378	0,173	16,240	4,399	8,813	0,880
RT-ColorAW	6	7,63	5,814	0,402	3,795	0,485	15,310	3,424	10,219	1,011
RealTimeABW	7	8,63	6,827	0,003	3,558	0,080	23,336	5,935	10,731	1,923
RealtimeBP	8	8,63	7,872	0,040	9,004	0,139	17,171	5,519	12,444	0,791
RealtimeVar	9	9,50	16,846	0,268	12,827	0,256	17,324	1,963	13,668	0,269
FastAggreg	10	10,25	6,061	0,049	6,433	3,845	20,214	4,815	11,935	2,168
RealtimeBFV	11	10,75	6,745	0,573	2,884	0,366	19,501	6,273	13,378	3,387
OptimizedDP	12	11,00	9,797	0,198	12,970	2,583	16,576	2,729	13,438	1,131
RTCensus	13	11,50	19,227	1,875	14,241	0,608	20,290	3,302	12,221	0,137
RealTimeGPU	14	11,75	10,640	0,102	20,266	0,508	17,573	3,318	16,462	1,510
ReliabilityDP	15	12,63	7,245	0,024	12,220	1,594	19,540	6,142	19,710	9,555
TwoWin	16	13,13	11,590	0,128	7,533	0,415	23,598	5,839	16,608	4,172
DCBGrid	17	13,63	21,045	2,471	11,186	0,590	22,173	6,067	14,876	0,692

- With regard to Table 5-8, different results are obtained for most of selected stereo methods (i.e. a difference of one or two positions on ranks for PlaneFitBP, CostFilter, GeoDif, DCBGrid, RealTimeABW, OptimizedDP, TwoWin, RTAdaptWgt, ReliabilityDP), and a more significant change for a few of them (i.e. a difference of three or more positions on ranks for RT-ColorAW, RealtimeVar, RealTimeBFV, RealTimeGPU, among others).

Table 5-9 and Table 5-10 show obtained evaluation results by selecting the *interior* and the *boundary* criterion, respectively. Obtained results shown in Table 5-9 and 5-10 keep a resemblance to results shown in Table 5-8.

Table 5-9 Evaluation of selected methods of near real-time and real-time performance under *interior* criterion, using BMP measure and Middlebury's evaluation model

Method	Rank	Avg.	Tsukuba	Venus	Teddy	Cones
			interior			
			BMP			
ADCensus	1	1,75	0,019	0,006	1,530	0,214
CostFilter	2	5,00	0,134	0,034	2,427	0,287
PlaneFitBP	3	5,75	0,003	0,050	3,595	1,034
RealtimeVar	4	6,25	0,268	0,256	1,963	0,269
RealtimeBP	5	7,25	0,040	0,139	5,519	0,791
GeoDif	6	7,75	0,579	0,175	3,233	0,281
RTAdaptWgt	7	7,75	0,057	0,173	4,399	0,880
RealTimeABW	8	8,25	0,003	0,080	5,935	1,923
RTCensus	9	9,25	1,875	0,608	3,302	0,137
RealTimeGPU	10	9,75	0,102	0,508	3,318	1,510
RT-ColorAW	11	10,25	0,402	0,485	3,424	1,011
OptimizedDP	12	10,50	0,198	2,583	2,729	1,131
FastAggreg	13	12,00	0,049	3,845	4,815	2,168
TwoWin	14	12,00	0,128	0,415	5,839	4,172
DCBGrid	15	12,75	2,471	0,590	6,067	0,692
ReliabilityDP	16	13,00	0,024	1,594	6,142	9,555
RealtimeBFV	17	13,75	0,573	0,366	6,273	3,387

Table 5-10 Evaluation of selected methods of near real-time and real-time performance under *boundary* criterion, using BMP measure and Middlebury's evaluation model

Method	Rank	Avg.	Tsukuba	Venus	Teddy	Cones
			boundary			
			BMP			
ADCensus	1	1,25	5,731	1,148	10,892	6,947
PlaneFitBP	2	3,00	5,256	1,708	14,712	10,604
CostFilter	3	4,75	7,606	2,419	16,043	7,663
GeoDif	4	5,00	7,638	3,017	13,291	8,087
RT-ColorAW	5	5,00	5,814	3,795	15,310	10,219
RTAdaptWgt	6	6,00	7,587	3,378	16,240	8,813
RealtimeBFV	7	7,75	6,745	2,884	19,501	13,378
FastAggreg	8	8,50	6,061	6,433	20,214	11,935
RealTimeABW	9	9,00	6,827	3,558	23,336	10,731
RealtimeBP	10	10,00	7,872	9,004	17,171	12,444
OptimizedDP	11	11,50	9,797	12,970	16,576	13,438
ReliabilityDP	12	12,25	7,245	12,220	19,540	19,710
RealtimeVar	13	12,75	16,846	12,827	17,324	13,668
RTCensus	14	13,75	19,227	14,241	20,290	12,221
RealTimeGPU	15	13,75	10,640	20,266	17,573	16,462
TwoWin	16	14,25	11,590	7,533	23,598	16,608
DCBGrid	17	14,50	21,045	11,186	22,173	14,876

The evaluation results obtained using the *all* criterion are illustrated in Table 5-11. The subsequent presented evaluation scenarios use this single criterion in order to focus the discussion on proposed evaluation methods, such as evaluation measures and models.

Table 5-11 Evaluation of selected methods of near real-time and real-time performance under *all* criterion, using BMP measure and Middlebury's evaluation model

Method	Rank	Avg.	Tsukuba	Venus	Teddy	Cones
			all			
			BMP			
ADCensus	1	1,00	1,485	0,254	6,216	7,254
CostFilter	2	2,75	1,847	0,391	11,796	8,245
PlaneFitBP	3	4,00	1,827	0,506	12,144	10,688
GeoDif	4	4,75	2,349	0,818	11,313	8,330
RTAdaptWgt	5	5,25	1,990	0,807	13,276	9,344
RealtimeBP	6	8,75	3,396	1,899	13,242	11,641
RealTimeABW	7	9,00	1,672	0,651	18,259	12,622
RT-ColorAW	8	9,25	3,075	1,715	13,953	11,869
RealtimeBFV	9	9,25	2,220	0,868	15,002	12,338
RealtimeVar	10	10,00	5,478	2,346	13,096	11,695
RTCensus	11	10,50	6,245	2,415	13,815	9,536
TwoWin	12	11,50	3,085	1,310	15,800	13,455
FastAggreg	13	12,00	2,114	4,749	15,165	12,603
OptimizedDP	14	13,50	3,782	4,742	13,889	13,737
RealTimeGPU	15	13,50	4,223	2,982	14,380	13,700
DCBGrid	16	13,50	7,259	1,910	17,234	11,947
ReliabilityDP	17	14,50	3,386	3,484	16,867	19,945

5.2.2 Selection of Evaluation Measures

As discussed in Section 4.2, the BMP does not consider estimation error magnitude, neither the inverse relation between depth and disparity. This issue can be alleviated using a measure or measures taking into account these factors. The MSE is a measure widely used, which in this context, considers estimation error magnitude. Table 5-12 shows evaluation results obtained by combining MSE and BMP under the *all* criterion. With regard to a combination of criteria it is convenient taking into account the proposed characterisation of evaluation measures, in particular the two last attributes: meaningful, and consistent. Thus, it should be highlighted that MSE and BMP have different meanings: the BMP is counting errors beyond a threshold, whilst the MSE provides information about squared differences of magnitude. Consequently, more changes can be expected when the MSE is used in isolation, with regard to be combined

with the BMP. Table 5-13 shows evaluation results under *all* criterion using the MSE. It can be observed that apart from the last ranked method, OptimizedDP, a new ranking is assigned to others methods. Moreover, method such as DCBGrid, RealTimeGPU, and RealtimeVar considerably improved in ranking position, whilst method such as RealtimeBP and RT-ColorAW considerably decreased in ranking position.

Table 5-12 Evaluation of selected methods of near real-time and real-time performance under all criterion by combining the MSE and the BMP measure, and using Middlebury's evaluation model

Method	Rank	Avg.	Tsukuba		Venus		Teddy		Cones	
			BMP	BMPRE	BMP	BMPRE	BMP	BMPRE	BMP	BMPRE
			all							
ADCensus	1	1,00	1,485	832,880	0,254	193,574	6,216	1856,230	7,254	2144,770
CostFilter	2	3,00	1,847	915,907	0,391	336,323	11,796	3467,500	8,245	2146,450
PlaneFitBP	3	4,75	1,827	1089,290	0,506	396,000	12,144	2478,180	10,688	4068,670
GeoDif	4	4,75	2,349	1248,260	0,818	510,294	11,313	3185,730	8,330	2424,650
RTAdaptWgt	5	5,13	1,990	1060,090	0,807	520,919	13,276	3120,560	9,344	2573,340
GradAdaptWgt	6	6,75	2,634	1363,850	1,392	1174,850	13,142	4151,580	7,674	2466,770
RealTimeABW	7	9,38	1,672	838,582	0,651	468,938	18,259	9776,230	12,622	6265,390
RealtimeBFV	8	9,50	2,220	1193,940	0,868	592,965	15,002	5701,760	12,338	4690,260
RealtimeVar	9	9,63	5,478	2116,930	2,346	1358,660	13,096	3550,020	11,695	2945,330
TwoWin	10	10,00	3,085	1780,570	1,310	965,369	15,800	4266,660	13,455	3867,600
RT-ColorAW	11	11,00	3,075	1827,690	1,715	1118,980	13,953	6009,230	11,869	4780,010
RealtimeBP	12	11,13	3,396	2527,320	1,899	1590,960	13,242	5794,210	11,641	4244,190
RTCensus	13	11,13	6,245	2637,120	2,415	1895,940	13,815	4698,150	9,536	2864,170
RealTimeGPU	14	12,75	4,223	2434,880	2,982	2225,300	14,380	4317,350	13,700	3960,630
DCBGrid	15	13,38	7,259	4639,970	1,910	1496,100	17,234	5992,270	11,947	3549,760
OptimizedDP	16	14,25	3,782	2755,500	4,742	3369,820	13,889	5410,520	13,737	5338,450
ReliabilityDP	17	15,50	3,386	2860,130	3,484	2341,250	16,867	6561,190	19,945	8722,740

Table 5-13 Evaluation of selected methods of near real-time and real-time performance under *all* criterion, using MSE measure and Middlebury's evaluation model

Method	Rank	Avg.	Tsukuba	Venus	Teddy	Cones
			all			
			MSE			
CostFilter	1	2,25	0,569	0,179	6,409	4,021
ADCensus	2	3,75	0,680	0,121	7,909	4,265
RTAdaptWgt	3	3,75	0,631	0,226	4,614	5,021
RealtimeVar	4	4,00	0,672	0,288	4,221	4,865
GeoDif	5	6,00	0,805	0,196	7,967	5,114
PlaneFitBP	6	6,25	0,708	0,242	3,384	13,604
RealtimeBFV	7	9,00	0,735	0,321	10,280	9,819
RealTimeGPU	8	9,00	1,161	0,474	4,503	7,663
RealTimeABW	9	10,00	0,605	0,241	31,923	34,618
TwoWin	10	11,00	1,077	0,326	11,804	12,219
RealtimeBP	11	11,25	1,275	0,559	7,354	12,456
DCBGrid	12	11,25	1,941	0,367	9,479	7,445
RT-ColorAW	13	11,75	0,976	0,339	10,940	22,024
RTCensus	14	11,75	1,179	0,426	15,648	6,509
FastAggreg	15	12,50	0,949	0,583	10,128	23,329
OptimizedDP	16	13,50	1,387	1,146	9,226	16,537
ReliabilityDP	17	16,00	1,457	0,765	12,849	44,717

However, neither the BMP nor the MSE consider the inverse relation between depth and disparity. The BMPRE considers both error magnitude and the inverse relation between depth and disparity. It has common properties with the BMP measure (i.e. are pondering the same estimation errors), and can be used, for instance, with already published data in order to allow a better analysis and deeper understanding of stereo methods behaviour. Obtained results for evaluating selected methods under the all criterion, using the BMP and the BMPRE, and the Middlebury's evaluation model are shown in Table 5-14. In this case, the evaluation is focused on the quantity of errors exceeding the used threshold, as well as on the magnitude and relation with depth of those errors. Moreover, the BMPRE measure is not limited to be used in conjunction with the BMP measure. The evaluation results obtained using exclusively the BMPRE measure, under the *all* criterion and the Middlebury's evaluation model, are shown in Table 5-15. In comparison to results shown in Table 5-11, the use of the BMPRE measure generates a significant redistribution of assigned ranks, as follows:

- the RTAdaptWgt, TwoWin, RealtimeVar, FastAggreg, RealtimeBFV, RTCensus, RealTimeGPU, DCBGrid stereo methods improve the ranking position,

- the PlaneFitBP, RealtimeBFV, RealTimeABW, RT-ColorAW, RealtimeBP, OptimizedDP stereo methods decrease the ranking position,
- whilst the remaining methods keep the ranking position: ADCensus, CostFilter, GeoDif, and ReliabilityDP.

Table 5-14 Evaluation of selected methods of near real-time and real-time performance under *all* criterion by combining the BMP and the BMPRE measure, and using Middlebury's evaluation model

Method	Rank	Avg.	Tsukuba		Venus		Teddy		Cones	
			BMP	BMPRE	BMP	BMPRE	BMP	BMPRE	BMP	BMPRE
			all							
ADCensus	1	1,00	1,485	832,880	0,254	193,574	6,216	1856,230	7,254	2144,770
CostFilter	2	3,00	1,847	915,907	0,391	336,323	11,796	3467,500	8,245	2146,450
PlaneFitBP	3	4,75	1,827	1089,290	0,506	396,000	12,144	2478,180	10,688	4068,670
GeoDif	4	4,75	2,349	1248,260	0,818	510,294	11,313	3185,730	8,330	2424,650
RTAdaptWgt	5	5,13	1,990	1060,090	0,807	520,919	13,276	3120,560	9,344	2573,340
GradAdaptWgt	6	6,75	2,634	1363,850	1,392	1174,850	13,142	4151,580	7,674	2466,770
RealTimeABW	7	9,38	1,672	838,582	0,651	468,938	18,259	9776,230	12,622	6265,390
RealtimeBFV	8	9,50	2,220	1193,940	0,868	592,965	15,002	5701,760	12,338	4690,260
RealtimeVar	9	9,63	5,478	2116,930	2,346	1358,660	13,096	3550,020	11,695	2945,330
TwoWin	10	10,00	3,085	1780,570	1,310	965,369	15,800	4266,660	13,455	3867,600
RT-ColorAW	11	11,00	3,075	1827,690	1,715	1118,980	13,953	6009,230	11,869	4780,010
RealtimeBP	12	11,13	3,396	2527,320	1,899	1590,960	13,242	5794,210	11,641	4244,190
RTCensus	13	11,13	6,245	2637,120	2,415	1895,940	13,815	4698,150	9,536	2864,170
RealTimeGPU	14	12,75	4,223	2434,880	2,982	2225,300	14,380	4317,350	13,700	3960,630
DCBGrid	15	13,38	7,259	4639,970	1,910	1496,100	17,234	5992,270	11,947	3549,760
OptimizedDP	16	14,25	3,782	2755,500	4,742	3369,820	13,889	5410,520	13,737	5338,450
ReliabilityDP	17	15,50	3,386	2860,130	3,484	2341,250	16,867	6561,190	19,945	8722,740

In this case, an improvement of ranking position implies that a stereo method is producing disparity estimation errors of a lower magnitude, and with a better inverse relation to depth, than disparity estimation errors produced by stereo method decreasing in their ranking position.

Table 5-15 Evaluation of selected methods of near real-time and real-time performance under *all* criterion based on the BMPRE measure, and using Middlebury's evaluation model

Method	Rank	Avg.	Tsukuba	Venus	Teddy	Cones
			all			
			BMPRE			
ADCensus	1	1,00	832,880	193,574	1856,230	2144,770
CostFilter	2	3,00	915,907	336,323	3467,500	2146,450
RTAdaptWgt	3	4,25	1060,090	520,919	3120,560	2573,340
GeoDif	4	4,75	1248,260	510,294	3185,730	2424,650
PlaneFitBP	5	5,00	1089,290	396,000	2478,180	4068,670
TwoWin	6	8,00	1780,570	965,369	4266,660	3867,600
RealtimeVar	7	8,25	2116,930	1358,660	3550,020	2945,330
RealtimeBFV	8	9,25	1193,940	592,965	5701,760	4690,260
RealTimeABW	9	9,75	838,582	468,938	9776,230	6265,390
RTCensus	10	10,25	2637,120	1895,940	4698,150	2864,170
RealTimeGPU	11	11,00	2434,880	2225,300	4317,350	3960,630
FastAggreg	12	11,50	1509,340	2019,860	4818,210	5288,430
RT-ColorAW	13	11,75	1827,690	1118,980	6009,230	4780,010
RealtimeBP	14	12,25	2527,320	1590,960	5794,210	4244,190
DCBGrid	15	12,25	4639,970	1496,100	5992,270	3549,760
OptimizedDP	16	14,50	2755,500	3369,820	5410,520	5338,450
ReliabilityDP	17	16,25	2860,130	2341,250	6561,190	8722,740

The BMPRE measure achieves a proper estimation of disparity estimation errors in terms of magnitude and position in relation to disparity ground-truth data, than the BMP measure. Nevertheless, it considers an error threshold, which, in practice, may turning it sensitive to a threshold selection by a user (i.e. evaluation results change as the used threshold change, but the specific 3D reconstruction achieved using the estimated disparity maps under evaluation remains fixed). In contrast, the SZE measure does not require any tolerance threshold. The evaluation of selected methods based on the SZE measure is shown in Table 5-16. The use of the SZE is aimed to evaluate the impact of estimation errors on 3D information recovering. It assumes that disparity ground-truth data is highly reliable. In comparison to evaluation results shown in Table 5-11, the use of the SZE produces the larger quantity on position ranking changes. These discrepancies in evaluation results are due to the theoretical differences of the SZE measure to other evaluation measures.

Table 5-16 Evaluation of selected methods of near real-time and real-time performance under all criterion based on the SZE measure, and using Middlebury's evaluation model

Method	Rank	Avg.	Tsukuba	Venus	Teddy	Cones
			all			
			SZE			
CostFilter	1	2,50	412,889	1085,280	232,257	113,226
RealtimeVar	2	3,50	675,465	1071,750	198,309	138,922
PlaneFitBP	3	5,25	512,201	1128,420	178,543	195,171
RTAdaptWgt	4	5,75	705,864	1131,720	246,920	144,835
RTCensus	5	6,00	813,622	840,636	536,328	129,387
GeoDif	6	7,75	1072,530	1099,380	290,026	130,222
ADCensus	7	8,25	994,842	862,403	443,840	162,017
RealTimeGPU	8	10,00	998,262	1312,780	257,529	172,864
TwoWin	9	10,00	1172,760	1164,990	255,662	183,417
RealTimeABW	10	10,25	432,096	1124,840	661,029	412,413
RealtimeBP	11	10,25	931,612	1223,870	311,724	198,708
FastAggreg	12	11,00	871,338	1297,210	309,994	276,044
DCBGrid	13	11,75	1225,310	811,411	1078,300	244,847
ReliabilityDP	14	12,00	865,506	1352,860	321,641	409,681
OptimizedDP	15	12,50	963,566	1423,960	366,830	211,007
RealtimeBFV	16	13,00	1024,620	1136,020	543,118	262,678
RT-ColorAW	17	13,25	1017,160	1183,840	356,573	440,515

5.2.3 Selection of the Evaluation Model

The proposed evaluation models are used in this section to handle the comparison of selected stereo methods as a decision making problem: giving a set of stereo methods under comparison, which of them shows a better trade-off according to considered evaluation elements and methods? In this regard, the A* and the $A^* - Groups$ evaluation models are based on multi-objective optimisation concepts and the Pareto Dominance relation. The proposed evaluation model aim to find a subset of elements in the solution space (i.e. the set of methods under comparison), according to vectors in the objective space (i.e. associated vectors of evaluation scores). Obtained evaluation results by applying the A* model, to selected stereo methods, under the *all* criterion and using the SZE measure are shown in Table 5-17. In this case, the interpretation of results is as follows: the set of methods composing the A* set are comparable among them, and, at the same time, more accurate than methods composing the set A'. Nevertheless, the A* evaluation model has a theoretical limitation: it does not provide enough information about the set A'. Moreover, the selection of a single method from the A* set may overload the judging capabilities of a decision maker.

Table 5-17 Evaluation results of methods with near real-time and real-time performance, under the all criterion, using the SZE measure, by applying the A* model.

Method	Set.	Tsukuba	Venus	Teddy	Cones
		all			
		SZE			
CostFilter	A*	412,889	1085,280	232,257	113,226
RealtimeVar	A*	675,465	1071,750	198,309	138,922
RTCensus	A*	813,622	840,636	536,328	129,387
PlaneFitBP	A*	512,201	1128,420	178,543	195,171
ADCensus	A*	994,842	862,403	443,840	162,017
DCBGrid	A*	1225,310	811,411	1078,300	244,847
GeoDif	A'	1072,530	1099,380	290,026	130,222
RTAdaptWgt	A'	705,864	1131,720	246,920	144,835
RealTimeABW	A'	432,096	1124,840	661,029	412,413
TwoWin	A'	1172,760	1164,990	255,662	183,417
RealTimeGPU	A'	998,262	1312,780	257,529	172,864
RealtimeBP	A'	931,612	1223,870	311,724	198,708
FastAggreg	A'	871,338	1297,210	309,994	276,044
RealtimeBFV	A'	1024,620	1136,020	543,118	262,678
ReliabilityDP	A'	865,506	1352,860	321,641	409,681
RT-ColorAW	A'	1017,160	1183,840	356,573	440,515
OptimizedDP	A'	963,566	1423,960	366,830	211,007

These drawbacks are alleviated in the $A^* - Groups$ model and by incorporating a method for reducing the cardinality of a Pareto set. Obtained evaluation results for selected methods, under the *all* criterion and using the SZE measure as well as the $A^* - Groups$ evaluation model are shown in Table 5-18. In this case, the interpretation of results involves the assigned group to the considered methods: the methods of the Group 1 are comparable among them and superior to rest of evaluated methods. The methods composing the Group 2 are comparable among them, and at the same time, superior to methods with a larger group number, and so on. In this way, the evaluation of stereo methods is performed in an extensive way, and providing information about each considered stereo method. On the other hand, it can be observed that the Group 1 is composed by 6 stereo methods. A selection of a method is performed based on the methods composing such group, as well on the different scores in evaluated criteria. In this regard, the method proposed in Section 4.3.3.1 for reducing the cardinality of a Pareto front brings support to the decision maker. The proposed method considers the functions u1 and u2. Computed values of function u1 and u2 for the methods composing the Group1 in Table 5-18, are shown in Table 5-19. It can be observed that the CostFilter method is achieving the best trade off (i.e. both associated values are lower

than the rest of computed function values). Consequently, and according to devised evaluation scenario (i.e. considering the impact on the 3D reconstruction of disparity estimation errors over the entire image produced by selected methods), this result can be interpreted as the CostFilter method should be selected.

Table 5-18 Evaluation results of methods with near real-time and real-time performance, under the all criterion, using the SZE measure, by applying the $A^* - Groups$ evaluation model

Method	Group	Tsukuba	Venus	Teddy	Cones
		all			
		SZE			
CostFilter	1	412,889	1085,280	232,257	113,226
RealtimeVar	1	675,465	1071,750	198,309	138,922
RTCensus	1	813,622	840,636	536,328	129,387
PlaneFitBP	1	512,201	1128,420	178,543	195,171
ADCensus	1	994,842	862,403	443,840	162,017
DCBGrid	1	1225,310	811,411	1078,300	244,847
GeoDif	2	1072,530	1099,380	290,026	130,222
RTAdaptWgt	2	705,864	1131,720	246,920	144,835
RealTimeABW	2	432,096	1124,840	661,029	412,413
TwoWin	3	1172,760	1164,990	255,662	183,417
RealTimeGPU	3	998,262	1312,780	257,529	172,864
RealtimeBP	3	931,612	1223,870	311,724	198,708
FastAggreg	3	871,338	1297,210	309,994	276,044
RealtimeBFV	3	1024,620	1136,020	543,118	262,678
ReliabilityDP	3	865,506	1352,860	321,641	409,681
RT-ColorAW	3	1017,160	1183,840	356,573	440,515
OptimizedDP	4	963,566	1423,960	366,830	211,007

Table 5-19 Values of functions u1 and u2 applied to stereo method composing group 1 under the *all* criterion and using the SZE measure.

Method	u1	u2
CostFilter	10	0,9236
RealtimeVar	12	1,3616
RTCensus	13	1,1059
PlaneFitBP	14	1,7448
ADCensus	16	1,5427
DCBGrid	19	3,0000

5.2.4 Evaluation in a Combination of Proposed Elements and Methods

Table 5-20 shows evaluation results for selected stereo methods under the proposed criteria (i.e. *boundary*, *interior* and *occluded*), using the SZE measure and the $A^* - Groups$ evaluation model.

Table 5-20 Evaluation results of methods with near real-time and real-time performance, under the boundary, interior, and occluded criteria, using the SZE measure, by applying the $A^* - Groups$ evaluation model

Method	Group	Tsukuba			Venus			Teddy			Cones		
		boundary	*interior*	*occluded*	*boundary*	*interior*	*occluded*	*boundary*	*interior*	*occluded*	*boundary*	*interior*	*occluded*
		SZE											
CostFilter	1	127,413	250,080	35,395	104,063	940,059	41,155	51,273	88,290	92,693	36,873	31,089	45,264
PlaneFitBP	1	128,413	316,315	67,474	97,084	979,196	52,144	33,764	81,731	63,049	51,258	37,967	105,946
GeoDif	1	166,705	857,880	47,949	99,396	946,666	53,315	55,239	151,773	83,015	44,063	31,764	54,394
RTAdaptWgt	1	151,026	500,022	54,816	109,260	968,997	53,466	47,688	138,570	60,662	45,858	37,153	61,824
ADCensus	1	159,581	793,167	42,094	80,744	744,691	36,967	162,180	196,600	85,061	46,793	30,565	84,660
GradAdaptWgt	1	132,289	256,801	41,162	114,265	1022,990	56,545	57,179	161,058	133,061	44,574	30,513	70,710
RealtimeVar	1	175,859	367,149	132,457	157,070	817,381	97,300	45,466	50,085	102,757	45,262	30,670	62,990
RealTimeABW	1	120,390	271,819	39,887	117,555	961,198	46,090	72,253	145,933	442,844	51,797	39,203	321,414
RTCensus	1	268,105	449,452	96,065	178,695	573,571	88,371	65,545	276,169	194,615	51,199	16,720	61,469
TwoWin	1	225,073	867,704	79,988	148,284	960,566	56,136	68,271	132,126	55,266	68,557	61,343	53,517
DCBGrid	1	333,614	778,270	113,423	137,662	609,233	64,516	154,132	786,496	137,674	88,849	23,830	132,168
RealtimeBP	2	187,957	560,109	183,546	158,732	955,403	109,739	64,745	140,237	106,743	56,760	36,133	105,815
RT-ColorAW	2	147,958	749,353	119,848	105,950	965,967	111,922	55,368	114,724	186,480	58,057	36,463	345,995
RealTimeGPU	2	181,857	669,270	147,135	207,590	997,908	107,281	53,527	93,254	110,748	64,267	37,608	70,989
RealtimeBFV	2	162,782	813,078	48,757	111,655	964,115	60,251	96,365	339,322	107,431	74,921	68,741	119,016
OptimizedDP	2	190,654	639,136	133,776	168,049	1106,330	149,586	68,555	144,581	153,694	59,492	35,783	115,732
ReliabilityDP	2	164,547	545,809	155,150	170,679	1046,590	135,589	68,613	109,006	144,022	79,413	81,249	249,019

Table 5-21 shows the values of functions u1 and u2 for the stereo methods composing the Group1, as it is shown in Table 5-20. Figure 5-7 is plotted based on the intermediated computed values of function u2 for the methods composing the Group 1 of evaluation results shown in Table 5-20. Figure 5-7 reflects the inherent multidimensionality of the obtained Pareto front, upon which a decision should be made about selecting a particular method.

Table 5-21 Values of functions u1 and u2 applied to stereo method composing Group 1 in Table 5- 20, under the boundary, interior, and occluded criteria and using the SZE measure

Method	u1	u2
CostFilter	35	1,4493
PlaneFitBP	57	2,2572
ADCensus	59	3,2418
GeoDif	60	3,0573
RTAdaptWgt	61	2,5801
RealtimeVar	72	3,3501
RealTimeABW	71	4,1480
RTCensus	77	3,8973
TwoWin	80	4,9297
DCBGrid	91	6,6913

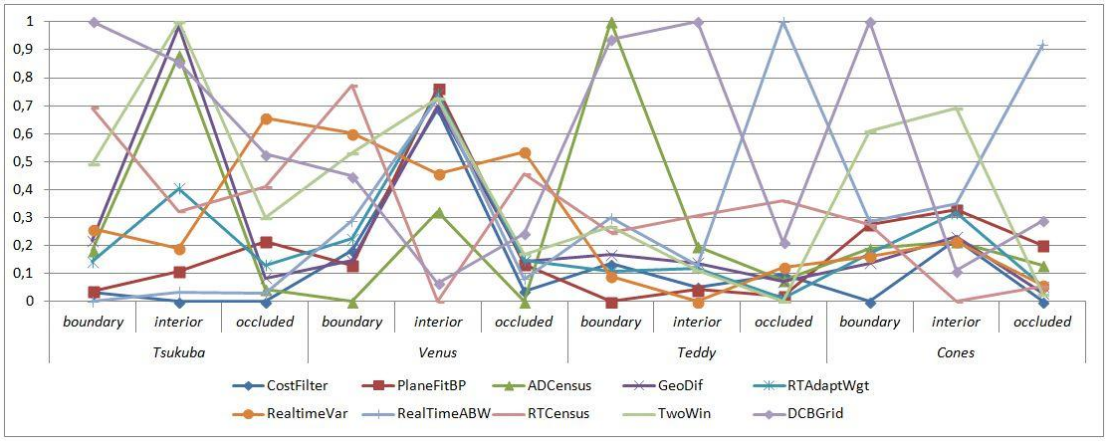


Figure 5-7 Intermediate computed values of function u2 for the methods composing the Group 1 shown in Table 5-20.

It can be observed in Table 5-21, that the lowest value in both functions is achieved by the CostFilter method. This result can be interpreted as that, among the different alternatives, and considering the different evaluation criteria the CostFilter stereo method should be selected (i.e. to be applied on imagery-test bed with similar conditions to the imagery-test bed used during the conducted evaluation process). The

intermediated computed values of function u2 for the method CostFilter are plotted in Figure 5-8, reflecting the reduced Pareto front obtained by applying the proposed method.

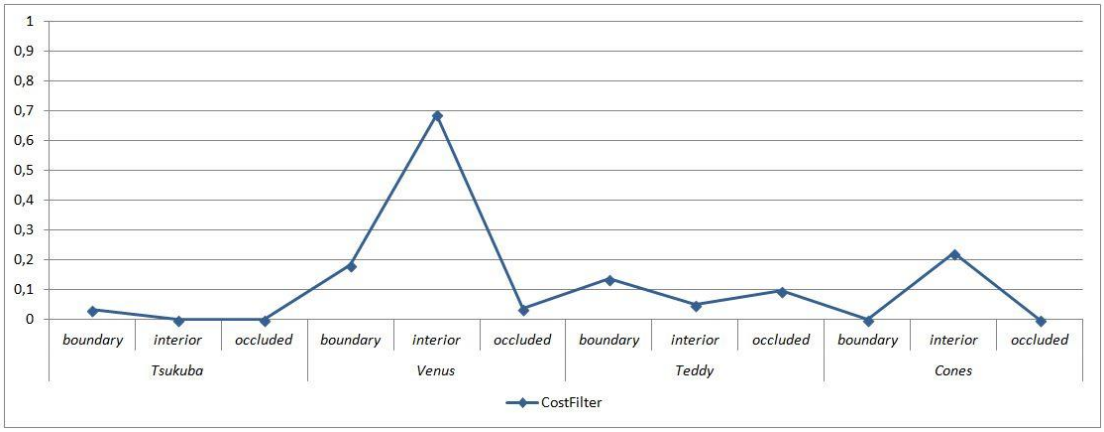


Figure 5-8 Intermediate computed values of function u2 for the CostFilter method.

5.3 Evaluation of Stereo Methods in Occluded Areas

In this section, the evaluation of stereo methods with regard to disparity estimations on occluded areas is addressed by the *occluded* criterion in the four stereo images of the benchmark. The interest for this evaluation scenario is twofold. Firstly, from a theoretical point of view disparity of occluded points cannot be estimated from image data. Secondly, in different application domains, dense maps are required. Consequently, it is expected that a stereo method, not only, produce accurate estimation at stereo visible areas, but also provide reasonable guesses at occluded regions (Sun et al., 2005). In practice, the disparity of occluded points has to be inferred from nearby stereo visible points. However, the presence of occluded points makes difficult the accurate disparity estimation. It is assumed that the goal in this evaluation scenario is to identify a set of methods obtaining accurate estimations at occluded regions.

5.3.1 Selection of Evaluation Elements and Methods

The SZE, which can be considered as the strictest measure among the different alternatives to be used, is selected. A set of 110 stereo methods from the Middlebury's repository, (excluding some reported but unpublished methods for which it is not

possible to known their building blocks) is selected for evaluation. The proposed A* Group evaluation model, and the Middlebury's evaluation model are both selected in order to compare obtained results. Figure 5-9 shows a compendium of evaluation results achieved by the $A^* - Groups$ model. A total of twelve groups were obtained.

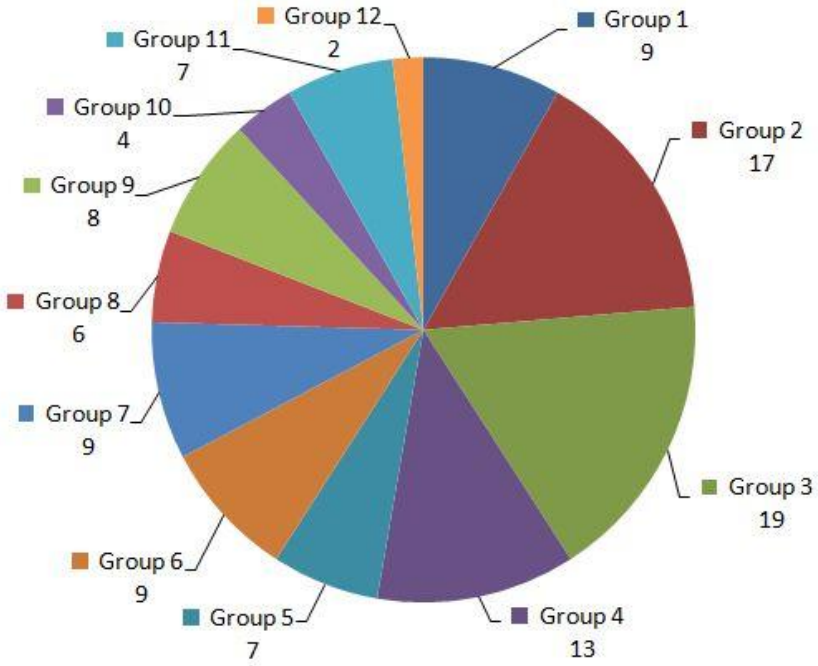


Figure 5-9 - Compendium of evaluation results obtained for diverse setreo methods of results

The methods composing the groups 1 and 2, according to the $A^* - Groups$ model are shown in Table 5-22. In comparison the top fifteen ranked methods according to the Middlebury's evaluation model are listed in Table 5-23. The number 15 is selected arbitrarily since in this model, the number of top performer methods is a free parameter. It can be observed, in the second column, that most of listed methods belong to the group 1 or 2 under the $A^* - Groups$ model.

With regard to the methods shown in Table 5-22, and composing the group 1, most of them are segmentation based methods. Interestingly, methods such as CurveletSupWgt (Mukherjee et al., 2010), and InteriorPtLP (Taylor & Bhusnurmat, 2008) are not segmentation based methods. For further information about these methods the reader is referred to the Section 3.3.1, and the Section 3.3.2, respectively.

Table 5-22 Evaluation results of stereo methods under the occluded criterion using the SZE measure and the $A^* - Groups$ evaluation model

Method	Group	Tsukuba	Venus	Teddy	Cones
		occluded			
		SZE			
WarpMat	1	25,727	30,217	30,098	42,585
SurfaceStereo	1	37,321	29,802	21,582	70,571
AdaptingBP	1	35,794	32,607	19,393	208,174
Segm+visib	1	26,686	43,206	24,072	60,875
Unsupervised	1	56,199	43,570	18,981	20,997
ObjectStereo	1	39,823	27,121	73,460	44,023
AdaptOvrSegBP	1	52,152	21,429	28,840	91,935
InteriorPtLP	1	36,078	62,848	32,749	37,275
CurveletSupWgt	1	44,450	36,074	33,614	40,696
OverSegmBP	2	36,772	38,921	32,892	49,945
OutlierConf	2	47,341	31,402	35,074	68,532
CostFilter	2	35,395	41,155	92,693	45,264
RegionTreeDP	2	30,931	52,088	34,111	79,531
GeoSup	2	35,918	35,480	53,897	70,399
CoopRegion	2	43,823	37,030	26,293	246,302
IterAdaptWgt	2	43,863	62,950	44,359	43,231
RDP	2	43,886	38,759	143,202	44,443
PatchMatch	2	30,400	31,818	382,693	211,895
Undr+OvrSeg	2	55,288	25,418	46,312	176,477
VSW	2	31,298	49,799	84,943	177,533
StereoSONN	2	65,670	34,101	137,294	58,373
AdaptWeight	2	47,381	59,388	47,560	47,824
GlobalGCP	2	119,846	56,289	31,706	61,227
DistinctSM	2	48,709	52,118	48,825	48,636
HistoAggr	2	32,971	46,089	157,460	95,079
AdaptDispCalib	2	38,997	34,198	232,874	252,912

Table 5-23 Top 15 ranked stereo methods under the occluded criterion using the SZE measure and the Middlebury's evaluation model

Method	A* group	Rank	Avg.	Tsukuba	Venus	Teddy	Cones
				occluded			
				SZE			
WarpMat	1	1	4,25	25,727	30,217	30,098	42,585
OverSegmBP	2	2	13,00	36,772	38,921	32,892	49,945
Segm+visib	1	3	13,75	26,686	43,206	24,072	60,875
CurveletSupWgt	1	4	13,75	44,450	36,074	33,614	40,696
SurfaceStereo	1	5	14,50	37,321	29,802	21,582	70,571
ObjectStereo	1	6	17,25	39,823	27,121	73,460	44,023
Unsupervised	1	7	18,50	56,199	43,570	18,981	20,997
InteriorPtLP	1	8	19,00	36,078	62,848	32,749	37,275
AdaptingBP	1	9	21,75	35,794	32,607	19,393	208,174
GeoSup	2	10	21,75	35,918	35,480	53,897	70,399
OutlierConf	2	11	22,00	47,341	31,402	35,074	68,532
CostFilter	2	12	22,50	35,395	41,155	92,693	45,264
AdaptOvrSegBP	1	13	23,00	52,152	21,429	28,840	91,935
RegionTreeDP	2	14	23,75	30,931	52,088	34,111	79,531
SymBP+occ	3	15	25,25	67,623	42,107	33,051	50,763

Table 5-24 Values of functions u1 and u2 applied to stereo method composing Group 1 in Table 5- 22, under the occluded criterion, using the SZE measure

Method	u1	u2
WarpMat	15	0,532
SurfaceStereo	18	0,895
AdaptingBP	19	1,608
Segm+visib	19	0,864
Unsupervised	19	1,535
ObjectStereo	22	1,723
AdaptOvrSegBP	22	1,427
InteriorPtLP	22	1,679
CurveletSupWgt	24	1,342

On the other hand, the proposal for reducing the cardinality of the Pareto front can be used in order to focus the attention into a reduced set of method. Table 5-24 shows computed values of function u1 and u2 for methods composing the group 1. It can be observed that the method WarpMat (Bleyer et al., 2009) achieves the lowest values in both functions.

5.4 Evaluation of Methods in Near and Far from Depth Discontinuities Areas

The evaluation scenario addressed in this section is motivated as a comparison of methods based on two aspects that all methods should address: the estimation of disparities at points near depth discontinuities (i.e. under the *boundary* criterion) and at stereo visible points far from depth discontinuities (i.e. under the interior criterion). In this way, is taken into account, and incoportaed into the evaluation process, that not all methods consider an occlusion model, on a non-trivial way to assign disparities in occluded regions. In this evaluation scenario, diverse stereo correspondence methods, regardlees the approaches on which they are based, are selected for comparison.

5.4.1 Selection of Evaluation Elements and Methods

The BMPRE measure (i.e. with the commonly used threshold of 1 pixel) is selected as the evaluation measure. A set of 110 stereo methods from the Middelbury's repository are selected for comparison. The proposed A* Group evaluation model, and the Middlebury's evaluation model are both selected in order to compare obtained results.

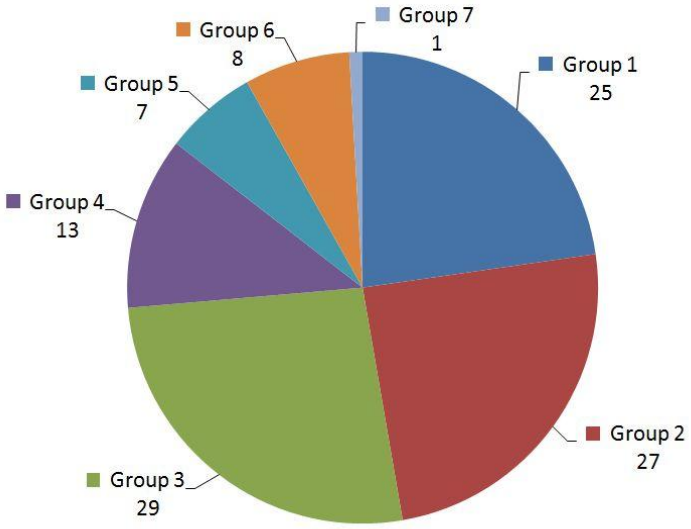


Figure 5-10 Composition of groups for the comparison of stereo methods under the *interior* and *boundary* criteria, using the $A^* - Groups$ model.

Table 5-25 Evaluation results of diverse stereo methods under the *boundary* and *interior* criteria, using the BMPRE measure and the $A^* - Groups$ model.

Method	Tsukuba		Venus		Teddy		Cones	
	boundary	interior	boundary	interior	boundary	interior	boundary	interior
	BMPRE							
DoubleBP	457,342	0,600	111,252	0,000	360,475	57,084	691,013	20,384
AdaptingBP	497,370	6,391	107,624	0,000	382,483	62,423	726,887	6,487
SubPixDoubleBP	499,915	24,271	113,239	0,000	371,209	43,666	703,896	20,353
CoopRegion	432,353	6,650	97,954	0,000	503,005	139,830	844,999	27,696
OutlierConf	437,987	0,800	130,249	1,319	574,552	128,925	632,001	57,647
ADCensus	516,730	7,800	72,205	0,590	644,829	345,522	705,154	7,751
SurfaceStereo	545,186	8,975	195,986	0,000	392,061	54,112	757,551	26,984
InfoPermeable	551,392	4,400	200,863	3,237	768,753	318,284	616,900	11,028
PlaneFitBP	468,690	1,200	88,286	6,721	497,919	261,489	1014,670	79,229
PatchMatch	750,091	178,675	127,189	8,017	483,776	60,058	649,636	15,857
ObjectStereo	603,038	16,800	392,010	74,340	491,010	167,886	538,041	9,202
Undr+OvrSeg	621,792	176,588	73,910	4,031	559,383	164,895	730,835	36,385
MVSegBP	477,603	11,200	78,153	7,731	560,283	219,511	1332,110	79,902
IterAdaptWgt	375,315	0,000	252,408	2,894	867,314	908,393	736,926	51,246
C-SemiGlob	757,435	158,998	256,198	2,836	567,620	237,947	722,631	8,782
RDP	420,829	58,600	122,784	39,956	792,502	538,179	725,382	19,510
OverSegmBP	662,102	43,975	215,745	63,648	656,446	317,568	669,690	25,035
AdaptOvrSegBP	484,877	288,533	71,754	7,157	764,875	572,232	937,616	67,592
CostFilter	617,365	90,800	152,378	9,417	1085,220	645,642	701,500	14,244
ASSM	508,215	35,975	374,889	97,696	929,134	499,720	616,627	61,339
RegionTreeDP	648,139	87,358	90,834	22,290	705,806	333,091	946,485	280,839
RealTimeABW	543,732	0,400	201,146	15,047	1649,730	1228,210	894,667	173,394
PlaneFitSGM	1323,820	172,187	990,314	23,699	1196,840	464,073	1207,780	4,878
CostRelax	2075,580	773,002	1300,650	72,648	1732,750	346,703	1400,870	1,789
RTCensus	1424,420	387,346	934,932	257,330	1394,480	1087,220	1183,660	6,472

Figure 5-10 shows a compendium of obtained results by the $A^* - Groups$ model. A total of eight groups were obtained. It can be observed that the three first groups involve the 74% of the compared methods. The methods composing the group 1 are listed in Table 5-25. It can be observed in Table 5-25, that the group 1 is composed by methods using different optimisation strategies (i.e. belief propagation, semi global matching, WTA, among others) and even by real-time performance methods. Moreover, it can be observed that several methods achieve a score of zero under the interior criterion for the Venus stereo image. This implies that there are methods which estimation errors do not exceed the considered trheshold. In general terms, the evaluation scores under the *boundary* criterion are higher that scores under the *interior* criterion for most of the methods.

Table 5-26 shows the values of the functions u1 and u2 for the methods composing the group 1 shown in Table 5-25. It can be observed that the DoubleBP (Yang et al., 2009) method achieves the lowest value in both functions.

Table 5-26 Values of functions u1 and u2 applied to stereo methods composing the group 1 in Table 5-25, under the *boundary* and *interior* criteria, using the BMPRE measure.

Method	u1	u2
DoubleBP	42	0,336
AdaptingBP	50	0,377
SubPixDoubleBP	58	0,405
CoopRegion	67	0,697
OutlierConf	69	0,628
ADCensus	73	0,773
SurfaceStereo	81	0,589
InfoPermeable	89	0,881
PlaneFitBP	90	1,210
PatchMatch	94	0,811
ObjectStereo	96	0,932
Undr+OvrSeg	101	0,986
MVSegBP	103	1,604
IterAdaptWgt	105	1,665
C-SemiGlob	108	1,146
RDP	110	1,312
OverSegmBP	121	1,273
AdaptOvrSegBP	122	1,905
CostFilter	125	1,632
ASSM	131	1,855
RegionTreeDP	135	2,345
RealTimeABW	140	3,231
PlaneFitSGM	151	3,372
CostRelax	165	5,538
RTCensus	168	5,220

Figure 5-10 reflects the Pareto front composed by the vector scores of methods in the group 1, using the intermediated computed values of the function u2, whilst Figure 5-11 reflects the reduction of the Pareto front by the proposed method.

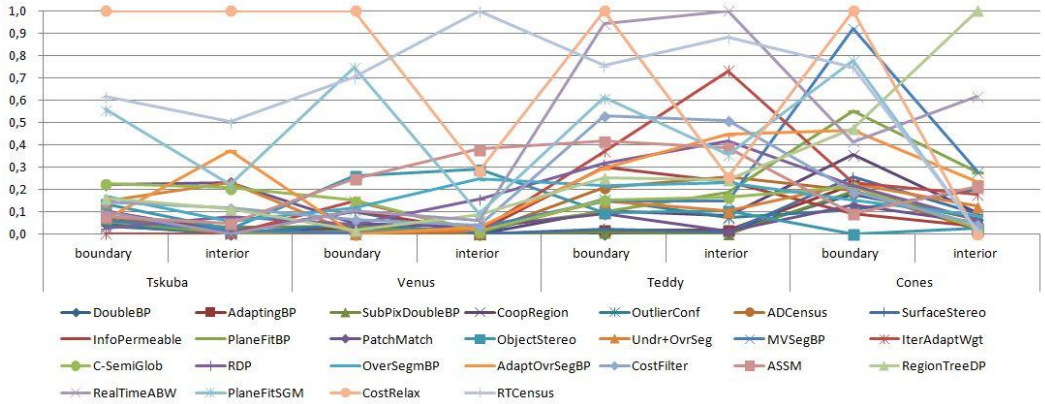


Figure 5-11 Intermediate computed values of function u2 for the methods composing the group 1 shown in Table 5-25.

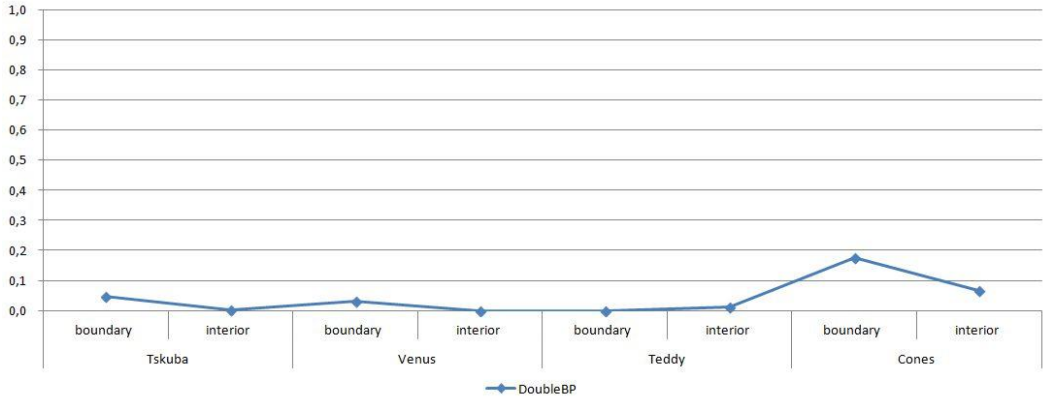


Figure 5-12 Intermediate computed values of function u2 for the method DoubleBP.

Obtained evaluation results using the Middlebury's evaluation model are shown in Table 5-27, which contains the top fifteen ranked. It can be observed that listed methods belong to the group 1 or to the group 2 by the $A^* - Groups$ evaluation model. In this way, the proposed model offers more detailed information on the performance of stereo methods. In addition, the DoubleBP method is the top ranked method.

Table 5-27 Evaluation Results using Midleburry's Evaluation Model

Method	A* group	Rank	Avg.	Tsukuba		Venus		Teddy		Cones	
				boundary	*interior*	*boundary*	*interior*	*boundary*	*interior*	*boundary*	*interior*
				BMPRE							
DoubleBP	1	1	6,88	457,342	0,600	111,252	0,000	360,475	57,084	691,013	20,384
AdaptingBP	1	2	8,50	497,370	6,391	107,624	0,000	382,483	62,423	726,887	6,487
SubPixDoubleBP	1	3	12,38	499,915	24,271	113,239	0,000	371,209	43,666	703,896	20,353
CoopRegion	1	4	12,50	432,353	6,650	97,954	0,000	503,005	139,830	844,999	27,696
OutlierConf	1	5	13,00	437,987	0,800	130,249	1,319	574,552	128,925	632,001	57,647
ADCensus	1	6	13,38	516,730	7,800	72,205	0,590	644,829	345,522	705,154	7,751
SurfaceStereo	1	7	16,25	545,186	8,975	195,986	0,000	392,061	54,112	757,551	26,984
WarpMat	1	8	18,38	508,632	13,117	175,737	0,958	432,573	110,247	752,941	55,901
InfoPermeable	1	9	19,00	551,392	4,400	200,863	3,237	768,753	318,284	616,900	11,028
PlaneFitBP	1	10	20,50	468,690	1,200	88,286	6,721	497,919	261,489	1.014,670	79,229
GlobalGCP	2	11	20,88	504,120	2,000	116,837	0,298	650,984	264,538	866,570	78,473
PatchMatch	1	12	23,63	750,091	178,675	127,189	8,017	483,776	60,058	649,636	15,857
SymBP+occ	2	13	24,75	461,497	7,900	188,949	0,172	634,888	141,305	1.038,590	144,851
Undr+OvrSeg	1	14	24,88	621,792	176,588	73,910	4,031	559,383	164,895	730,835	36,385
GeoSup	2	15	24,88	674,756	12,467	115,761	1,044	669,853	370,654	822,491	29,364

5.5 Chapter Summary

- The quantity of badly estimated matches according to the BMP measure, (i.e. estimated disparities beyond one pixel of error tolerance) associated to the *interior* criterion, on the Tsukuba and the Venus images are by several stereo methods are considerably low, and even zero in several cases. These values may be indicating that such images are no longer challenging for the state-of-the-art in stereo correspondence methods.
- The evaluation in occluded areas shows that there are methods able to assign accurate disparity values in these regions. Most of these methods use a segmentation-based approach.
- The evaluation of diverse stereo methods under the *interior* and the *boundary* criteria using the BMPRE measure, shown that the estimation of disparities in areas near to depth discontunities or occluded areas is the most challenging phenomenon for a stereo method (i.e. at least under the used test-bed), even in the presence of textureless regions and repetitive patterns. In addition, a group of level 1, composed by both global and local methods, as well as by real-time performance methods was obtained using the evaluation elements and methods mentioned above in conjunction with the $A^* - Groups$ evaluation model.
- The Middlebury's evaluation model is more suited to be used in a competition than in an objective comparison of stereo methods. In this regard, proposed evaluation models can be used for properly comparing different stereo methods. The obtained evaluation results by the proposed models can be used for selecting a particular stereo method under a determined evaluation scenario, as well as for identifying algorithmic modules producing good results under a specific evaluation scenario.

CHAPTER 6.

FINAL REMARKS AND FUTURE WORK

Chapter Contents

6.1 Discussion

A research question was addressed in the presented research:

- Which are the method or methods accurately matching corresponding points, in order to allow a better 3D information recovery in terms of depth calculations, among a set of stereo correspondence methods being compared, under an specific evaluation scenario?.

In this regard the following concerns can be identified:

- The selection of an imagery test-bed is indeed a very relevant element in an evaluation process. Aspects such as image capturing conditions, image realism, image content, the reliability of disparity ground-truth data, and the relation of image content to some application domain, among others, have to be considered for the selection of the test-bed.
- A proper use of evaluation criteria makes possible establishing a relation between the presence of mismatches and stereo image phenomena. The composition of each criterion is guided according to the phenomenon of interest, whilst the selection of criteria may be based on an application domain, which may introduce different relevancies to different criteria.

- An assessment of disparity maps for evaluating the impact of mismatches on depth calculation mainly relies on the evaluation measures. Thus, an evaluation measure should incorporate into its formulation such requirement. In this regard, two basic aspects were indentified in the thesis: the magnitude of the disparity estimation error, and the true distance between the point being evaluated and the stereo camera system.
- A trivial comparison among stereo methods does not exist, since any type of comparison in this matter requires an analisys on both isolated and comparative performance. Moreover, misinterpretations on obtained evaluation results may cause a misunderstanding on the state-of-the-art on stereo correspondence methods.

6.2 Remarks on Obtained Evaluation Results

The discussion in the literature on evaluation methodologies for stereo correspondence methods has been mostly focused on a single evaluation element: the selection of an imagery test-bed. However, this is not the only evaluation element or methods deserving attention during an evaluation process, since, as it was illustrated in the experimental evaluation, the selection of any evaluation element or method has an impact on obtained evaluation results.

The experimentation under the *near*, the *mid* and the *far* evaluation criteria did not shown conclusive results due to the short depth range in used imagery test-bed.

The *all* criterion is in fact an absence of evaluation criteria since it does not relate detected errors to any defined stereo image phenomenon.

A very low score of the BMPRE measure (considering a threshold of 1 pixel) for the Tsukuba and the Venus images under the proposed *interior* criterion, was shown by several stereo methods. This score may be interpreted as an evidence of these two stereo images are not longer challenging for state-of-the-art methods.

The evaluation of several stereo methods under the proposed criteria of *interior* and *boundary*, using different evaluation measures and the proposed evaluation model showed, that there is not a single optimisation strategy which can be considered as the

one producing the best results, as well as there are also few local methods with comparable performance to global ones.

Disparity estimation on areas near depth discontinuities and occluded regions (i.e. regions associated to the *boundary* criterion) is, according to used test-bed imagery, more challenging than disparity estimation on smooth surfaces far enough from discontinuities (i.e. regions associated to the *interior* criterion).

6.3 Summary of Contributions

A set of specific contributions are pointed out below. The relevance of each one of them with regard to state-of-the-art is briefly highlighted.

A theoretical foundation for evaluation criteria, based on sets partitions is presented. The formulation of evaluation criteria based on sets partitions avoids multiple inclusions of points under different criteria, allowing a clear interpretation of obtained evaluation scores. Moreover, the proposed criteria involve the evaluation of disparity estimation on occluded regions. In contrast, in conventional evaluation methodologies, evaluation criteria are used empirically as binary segmentations. In this way, a point can be included in more than one criterion. This cause a bias on gathered scores, as well as on interpreting obtained evaluation results, since it is not possible to properly relate computed errors to image phenomena, nor vice versa.

Two evaluation measures are formulated in order to assess the impact of mismatches on depth calculations: the SZE and the BMPRE. On the one hand, the formulation of the SZE is inherently related to the depth calculation in a stereo system. In this way it considers disparity estimation error magnitude and the inverse relation between depth and disparity. It is suited to be used with highly reliable disparity ground-truth data, and does not consider an error threshold. It requires information about the stereo camera system setup, but if it is not available, computed error scores are up to a factor of 3D error. On the other hand, the BMPRE measure considers an error threshold, which can be fixed according to application domain, and even, based on the reliability of disparity ground-truth data. It considers the relative error caused by a mismatch, with regard to magnitude and true disparity. It can be used in conjunction with a conventional evaluation measure, or in isolation in order to properly assess disparity maps accuracy.

Although conventionally different evaluation measures have been used for comparing estimated disparity maps against disparity ground-truth data none of them consider that disparity estimation is an intermediate step in the 3D information recovery process, neither the inverse relation between depth and disparity, nor the inherent depth error of stereo camera systems.

The A* and the $A^* - Groups$ evaluation models are presented. The A* evaluation model addresses the evaluation of disparity maps as multi-objective optimisation problem based on the Pareto Dominance relation. In this way, it computes a Pareto front composed by incomparable evaluation vectors, associated to stereo methods of comparable behaviour among them. Moreover, it considers a formulation for interpreting evaluation results based on the cardinality of the obtained Pareto front set. In this way, a subjective interpretation of objective evaluation results is avoided. The A* evaluation model is extended by the $A^* - Groups$ evaluation model in order to perform an exhaustive comparison of all the methods considered during the evaluation. The extension incorporates a grouping algorithm associating each compared stereo correspondence method with a label identifying each group of comparable performance. In contrast, conventional evaluation models proceed by operating different evaluation scores among them in order to compute a single value as indicative of methods performance. However, most of the operated values are, by definition incommensurable among them. Moreover, conventionally used models do not consider a formulation for interpreting evaluation results. Thus different and contradictories interpretations of the same results may arise.

A method for reducing the cardinality of the Pareto front is introduced. The proposed evaluation models compute a Pareto front set based on evaluation score vectors. From such set, as in any other optimisation problem addressed by the Pareto dominance relation, a solution has to be selected by a decision maker. It is quite common that the cardinality of the obtained set, as well as the multidimensionality of the problem overload the judging capacity of the decision maker. The proposed method is not only useful in the context of comparing stereo methods, but is also of general purpose, and can be used in the decision making stage of any multi-objective optimisation problem. The novelty of the proposed method consist in selecting of a solution from a Pareto front set as a multi-objective problem based on two functions

computed on the decision space of the original problem. In contrast, conventional approaches for handling this decision making problem require the specification of preferences or additional information by the decision maker in order to guide the selection of a solution. However such preferences are not always available or may do not even exist.

6.4 Future Work

Three future work directions are identified.

- To apply the proposed evaluation model into a different imagery test-bed (i.e. such as the stereo Kitti benchmark data (Geiger et al., 2012) or the augmented Lauven data set (Ladicky et al., 2010)), aiming the incorporation into the evaluation process of some aspects or indicators about the task for which the disparity map is being estimated (i.e. evaluating achieved goals under a specific application domain based on the estimated disparity map).
- To generate disparity ground-truth data, related to realistic image capturing conditions (i.e. using stereo camera systems under poorly controlled and/or uncontrolled capturing conditions), involving a similar content in an intra-set level, as well as different image content in an inter-set level, aiming being of interest to multiple application domains. In particular, a semi-automatic process on which, in a first step, an initial disparity map is generated by the consensus of multiple stereo methods, and later in a second step, the map is corrected and refined by human supervision, will be the first approach to be explored.
- To extend the proposed methodology in order to cover, not only stereo images, but also stereo sequences.

References

A

(Ahuja & Tuceryan, 1989) Ahuja N. and Tuceryan M., “Extraction of Early Perceptual Structure in Dot Patterns: Integrating Region, Boundary, and Component Gestalt”, Computer Vision, Graphics, and Image Processing, vol. 48, pp. 304-356, 1989.

(Ayache, 1991) Ayache N., “Artificial vision for mobile robots: stereo vision and multisensory perception”, MIT Press, 1991.

(Awrangjeb et al., 2012) Awrangjeb M., Lu G., Fraser C.S., “Performance Comparisons of Contour-Based Corner Detectors”, Image Processing, IEEE Transactions on , vol.21, no.9, pp.4167-4179, 2012.

B

(Baker & Szeliski, 1998) Baker S., Szeliski R., and Anandan P. “A layered approach to stereo reconstruction”. In CVPR, pp. 434–441, 1998.

(Ballard & Brown, 1982) Ballard D. and Brown C., “Computer Vision Department of Computer Science”, University of Rochester, New York ISBN : 0-13-165316-4 Prentice Hall, 1982.

(Barnard & Fischler, 1982) Barnard S. and Fischler M., “Computational stereo”, ACM Computing Surveys, vol. 14, no. 4, pp. 553–572, 1982.

(Barron et al., 1994) Barron J., Fleet D., and Beauchemin S., “Performance of optical flow techniques”. IJCV, 12(1):43–77, 1994.

(Bay et al., 2008) H. Bay, T. Tuytelaars, and L. Van Gool, “Speeded-Up Robust Features(SURF)”, Computer Vision and Image Understanding, vol. 110, pp. 346–359, 2008.

(Beaudet, 1978) Beaudet, P.R., “Rotational invariant image operators”, 4th Intern. Conf Patt. Recog., pp. 579-583, 1978.

(Ben Said et al., 2010) Ben Said, L., Bechikn, S., Ghédira, K, “ The r-Dominance: A New Dominance Relation for Interactive Evolution Multi-criteria ision Making. IEEE Trans. on Evolutionary Computation 14(5), 801–818 (2010)

(Bentley & Wakefield, 1997) Bentley, P., Wakefield, J, “Finding Acceptable solutions in the Pareto-Optimal Range using Multiobjective Genetic Algorithms. In: P. Chawdhry, et al. (Eds), Soft Computing in engineering Design and Manufacturing, pp. 231-240, 1997.

(Birchfield & Tomasi, 1998) Birchfield, S., and Tomasi, C., “Depth discontinuities by pixel-to-pixel stereo”, Computer Vision, Sixth International Conference on , pp.1073-1080, 1998.

(Blanco et al., 2009) Blanco J., Moreno F., and Gonzalez J., “A collection of outdoor robotic datasets with centimeter-accuracy ground truth”, Journal of Autonomous Robots, Springer, pp. 327 - 351, 2009.

(Bleyer & Gelautz, 2004) Bleyer, M., Gelautz, M., “A layered stereo algorithm using image segmentation and global visibility constraints”, Image Processing, 2004. ICIP. International Conference on , vol.5, pp. 2997- 3000, 2004.

(Bleyer & Gelautz, 2008) M. Bleyer and M. Gelautz. Simple but effective tree structures for dynamic programming-based stereo matching. In VISAPP, vol. 2, pp. 415–422, 2008.

(Bleyer et al., 2009) Bleyer, M., Gelautz, M., Rother, C., Rhemann, C., , “A stereo approach that handles the matting problem via image warping”, Computer Vision and Pattern Recognition - CVPR IEEE Conference on, pp. 501-508, 2009.

(Bleyer et al., 2010) Bleyer, M., Rother, C., Kohli, P., “Surface stereo with soft segmentation”, Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on , pp.1570-1577, 2010.

(Bleyer et al., 2011) Bleyer, M., Rother, C., Kohli, P., Scharstein, D., Sinha, S., “Object stereo — Joint stereo matching and object segmentation”, Computer Vision and Pattern Recognition - CVPR, IEEE Conference on , pp. 3081-3088, 2011.

(Blostein & Huang, 1987) Blostein S., and Huang T., “Error Analysis in Stereo Determination of 3-D Point Positions”, Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.PAMI-9, no.6, pp.752-765, Nov. 1987

(Brockhoff et al., 2006) Brockhoff D., Zitzler E., “Are All Objectives Necessary? On Dimensionality Reduction in Evolutionary Multiobjective Optimization”, In: Runarsson, T.P., Beyer, H.-G., Burke, E.K., Merelo-Guervós, J.J., Whitley, L.D., Yao, X. (Eds.) PPSN 2006. LNCS, Springer, Heidelberg, vol. 4193, pp. 533–542. 2006.

(Brown et al, 2003) Brown M., Burschka D., and Hager G., “Advances in computational stereo”, IEEE Trans. Pattern Anal. Machine Intell., vol. 25, no. 8, pp. 993–1008, 2003.

(Brunet et al., 2012) Brunet D., Vrscay E., Wang Z., “On the Mathematical Properties of the Structural Similarity Index”, Image Processing, IEEE Transactions on , vol.21, no.4, pp.1488-1499, 2012.

(Bolles et al., 1993) Bolles R., Baker H., and Hannah M., “The JISCT Stereo Evaluation”, In ARPA Image Understanding Workshop, pp. 263-274, 1993.

(Borgefors, 1986) Borgefors G., “Distance transformations in digital images”, Computer Vision, Graphics and Image Processing, 34(3):344–371, 1986.

(Boyd & VandenBerghe, 2004) Boyd S., and VandenBerghe L., “Convex Optimization”, Cambridge University Press, 2004.

(Boyer & Kak, 1987) Boyer, K., and Kak A., “Color-Encoded Structured Light for Rapid Active Ranging”, Pattern Analysis and Machine Intelligence, IEEE Transactions on - PAMI, vol. 9, no.1, pp.14-28, 1987.

(Boykov et al., 1998) Boykov Y., Veksler O., and R. Zabih, “A Variable Window Approach to Early Vision”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1283-1294, 1998.

(Boykov et al., 1999) Boykov Y., Veksler O., and Zabih R., “Fast approximate energy minimization via graph cuts”, In International Conference on Computer Vision, pp. 377–384, 1999.

C

(Cabezas et al., 2011) Cabezas I., Padilla V., and Trujillo M., “A Measure for Accuracy Disparity Maps Evaluation”, LNCS 7042, Springer, pp. 223-231, 2011.

(Cabezas & Trujillo 2011) Cabezas I., and Trujillo M., “A Non-linear Quantitative Evaluation Approach for Disparity Estimation - Pareto Dominance Applied in Stereo Vision”, VISAPP, SciTePress, pp.704-709, 2011.

(Cabezas et al., 2012 a) Cabezas I., Trujillo M., and Florian M. “An Evaluation Methodology for Stereo Correspondence Algorithms”, In Proceedings of the International Conference on Computer Vision Theory and Applications, pp. 154-163, 2012.

(Cabezas et al., 2012b) Cabezas I. Padilla V., Trujillo M., and Florian M., “On the impact of the error measure selection in evaluating disparity maps”, In World Automation Congress (WAC), 2012 , pp.1-6, 2012.

(Cabezas & Trujillo, 2012) Cabezas I., and Trujillo M., “A Method for Reducing the Cardinality of the Pareto Front “ LNCS 7441 pp. 829-836, 2012.

(Cabezas et al., 2012d) Cabezas I., Padilla V., and Trujillo M., “BMPRE: An Error Measure for Evaluating Disparity Maps”, In Proc. International Conference on Signal Processing. Vol 2, pp. 1051- 1055 , 2012.

(Cabezas & Trujillo, 2013) Cabezas I., and Trujillo M., “Methodologies for Evaluating Disparity Estimation Algorithms, Robotic Vision”, Technologies for Machine Learning and Vision Applications, Jose Garcia-Rodriguez and Miguel A. Cazorla Quevedo Eds., IGI-Global, pp.154-172, 2013.

(Canny,1986) Canny J., “A Computational Approach to Edge Detection”, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.PAMI-8, no.6, pp.679-698, 1986.

(Carneiro & Jepson, 2002) Carneiro G., and Jepson A., “Phase-Based Local Features”, Proc. Seventh European Conf. Computer Vision, pp. 282-296, 2002.

(Carneiro & Jepson, 2003) Carneiro G., and Jepson A., “Multi-scale phase-based local features”, Computer Vision and Pattern Recognition, 2003. In Proceedings IEEE Computer Society Conference on , vol.1, pp. I-736- I-743, 18-20 2003.

(Cech et al., 2011) Cech, J., Sanchez-Riera J., and Horaud R., “Scene flow estimation by growing correspondence seeds”, Computer Vision and Pattern Recognition (CVPR), IEEE Conference on , , pp.3129-3136, 2011.

(Chang et al., 2011) Xuefeng Chang, Zhong Zhou, Liang Wang, Yingjie Shi, and Qinping Zhao, “Real-Time Accurate Stereo Matching Using Modified Two-Pass Aggregation and Winner-Take-All Guided Dynamic Programming”, 3D Imaging,

Modeling, Processing, Visualization and Transmission - 3DIMPVT, International Conference on , pp.73-79, 2011.

(Chandler & Hemami, 2007) Chandler D., and Hemami S., “VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images”, Image Processing, IEEE Transactions on , vol.16, no.9, pp.2284-2298, 2007.

(Chen & Medioni, 1998) Chen Q., and Medioni G., “Building human face models from two images”, Multimedia Signal Processing, 1998 IEEE Second Workshop on , pp.117-122, 1998.

(Chen et al., 2008) Chen S., Li Y., and Jianwei Zhang, “Vision Processing for Realtime 3-D Data Acquisition Based on Coded Structured Light”, Image Processing, IEEE Transactions on , vol.17, no.2, pp.167-176, 2008.

(Cheng, 1995) Yizong Cheng, “Mean shift, mode seeking, and clustering”, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.17, no.8, pp.790-799, 1995.

(Christoudias et al., 2002) Christoudias C., Georgescu B., and Meer P., “Synergism in low level vision”, Pattern Recognition, 2002. Proceedings. 16th International Conference on , vol.4, no., pp. 150- 155 vol.4, 2002.

(Coleman et al., 2007) Coleman S., Kerr D., and Scotney B., “Concurrent Edge and Corner Detection”, Image Processing, 2007. ICIP 2007. IEEE International Conference on, vol 5, pp. 273 – 276, 2007.

(Comaniciu & Meer, 1997) Comaniciu D., and Meer P., “Robust analysis of feature spaces: color image segmentation”, in Proc. of IEEE conference on Computer Vision and Pattern Recognition, pp. 750-755, 1997.

(Comaniciu & Meer, 2002) Comaniciu D., and Meer P., “Mean shift: A robust approach toward feature space analysis”, IEEE:PAMI, 24(5):603–619, 2002.

(Courtney et al., 1997) Courtney P., Thacker N., and Clark A., “Algorithmic modeling for performance evaluation”, In I Conference on Machine Vision Applications (MVA), pp. 219–228, 1997.

(Cox et al.,1996) Cox I., Hingorani S., Rao S., and Maggs B., “A Maximum Likelihood Stereo Algorithm”, Computer Vision, Graphics, and Image Processing, vol. 63, no. 3, pp. 542-567, 1996.

(Cvetkovic & Parmee, 2002) Cvetkovic D., and Parmee I., “Preferences and their Application in Evolutionary Multiobjective Optimisation”, IEEE Trans. Evolutionary Computation 6(1), 42–57, 2002.

(Cvetkovic & Coello, 2005) Cvetkovic D., and Coello C, “Human Preferences and their Applications in Evolutionary Multi-objective Optimisation”, In: Yaochun, J. (Eds). Knowledge Incorporation in Evolutionary Computation, Springer , pp. 479-502, 2005.

D

(Darrel,1998) Darrel T., “A Radial Cumulative Similarity Transform for Robust Image Correspondence”, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 656-662, 1998.

(Das, 1999) Das I, “On Characterizing the ‘knee’ of the Pareto Curve Based on normal-Boundary Intersection”, In Structural and Multidisciplinary Optimization, 18(2): pp. 107-115, 1999.

(Davis et al., 2005) Davis J., Nehab D., Ramamoorthi R., and Rusinkiewicz S., “Spacetime stereo: a unifying framework for depth from triangulation”, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.27, no.2, pp. 296-302, 2005.

(Deb et al.,2002) Deb K., Pratap A., Agarwal S., and Meyarivan T., “A fast and elitist multiobjective genetic algorithm: NSGA-II”, Evolutionary Computation, IEEE Transactions on , vol.6, no.2, pp.182-197, 2002.

(De-Maeztu et al., 2011) De-Maeztu L., Villanueva A., and Cabeza R., “Stereo matching using gradient similarity and locally adaptive support-weight”, Pattern Recognition Letters, Vol. 32, No. 13, pp. 1643-1651, 2011.

(De-Maeztu et al., 2012) De-Maeztu L., Villanueva A., and Cabeza, R., “Near Real-Time Stereo Matching Using Geodesic Diffusion”, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.34, no.2, pp. 410-416, 2012.

(Deriche & Giraudon, 1990) Deriche R., and Giraudon G., “Accurate Corner Detection: An Analytical Study”, Proc of Int. Conf. Computer Vision pp. 66-70, 1990.

(Deriche & Giraudon, 1993) Deriche R., and Giraudon G.,. “A computational approach for corner and vertex detection”, International Journal of Computer Vision 10 (2), 101-124, 1993.

(Dhond & Aggarwal, 1989) Dhond U., and Aggarwal J., “Structure from stereo—a review”, IEEE Trans. Syst., Man, Cybern., vol. 19, no. 6, pp. 1489–1510, 1989.

(Dreschler & Nagel, 1982) Dreschler L, and Nagel H., “On the selection of critical points and local curvature extrema of region boundaries for interframe matching”, Intern. Conf. Patt. Recog., pp. 542-544, 1982.

(Dongbo et al., 2010) Dongbo Min, Sehoon Yea, and Vetro A., “Temporally consistent stereo matching using coherence function”, 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), pp.1-4, 2010.

E

(Egnal et al., 2004) Egnal G., Mintz M., and Wildes R, “A stereo confidence metric using single view imagery with comparison to five alternative approaches”, Image Vision Computing 22, 943–957, 2004.

F

(Faugeras,1992) Faugeras O., “What can be seen in three dimensions with an uncalibrated stereo rig?”, In Proc. 2nd European Conference on Computer Vision, G. Sandini (Ed.), Santa Margherita Ligure, Italy, Springer-Verlag, vol. LNCS 588, pp. 563–578.

(Faugeras, 1993) Faugeras O., “Three-Dimensional Computer Vision: A Geometric Viewpoint”, MIT Press, 1993.

(Felzenszwalb & Huttenlocher, 2004) Felzenszwalb P., and Huttenlocher D., “Efficient belief propagation for early vision”, In CVPR, pp. 261–268, 2004.

(Forsyth & Ponce 2011) Forsyth D., and Ponce J., “Computer Vision: A Modern Approach”, Second Edition Pearson Education, Prentice Hall, 2011.

(Freeman & Adelson, 1991) Freeman W., and Adelson E., “The design and use of steerable filters”, PAMI, 13(9):891–906, 1991.

(Förstner & Gülch, 1987) Förstner W., and Gülch E. “A fast operator for detection and precise location of distinct points, corners and centres of circular features”, In Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281–305, 1987.

(Fua, 1993) Fua P., “A parallel stereo algorithm that produces dense depth maps and preserves image features”, Machine Vision and Applications, vol. 6, pp. 35–49, 1993.

(Fusiello & Trucco, 1997) Fusiello A., Roberto V., and. Trucco E., “Efficient stereo with multiple windowing”, In Proceedings CVPR, pp. 858-863, 1997,

(Fusiello et al., 2000) Fusiello A., Trucco E., and Verri A., “A compact algorithm for rectification of stereo pairs”, Machine Vision and Applications, Vol 12 No. 1, pp. 16-22, 2000.

(Förstner, 1997) Förstner W., “10 pros and cons against performance characterization of vision algorithms”, Machine Vision Applications 9, pp. 215–218, 1997.

G

(Gallup et al., 2008) Gallup D., Frahm J., Mordohai P., and Pollefeys M., “Variable baseline/resolution stereo”, Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on , pp.1-8, 23-28, 2008.

(Geiger et al., 1995) Geiger D., Ladendorf B., and Yuille A., “Occlusions and binocular stereo”, International Journal of Computer Vision, Vol 14, pp. 211–226, 1995.

(Geiger et al., 2012) Geiger A., Lenz P., and Urtasun R., “Are we ready for autonomous driving? The KITTI vision benchmark suite”, Computer Vision and Pattern Recognition - CVPR, IEEE Conference on , pp. 3354-3361, 1 2012.

(Gerrits & Bekaert, 2006) Gerrits M., and Bekaert P., “Local stereo matching with segmentation-based outlier rejection”, in CRV, 2006.

(Gimel'farb, 2001) Gimel'farb G., "Binocular Stereo by Maximizing the Likelihood Ratio Relative to a Random Terrain", Robot Vision, Lecture Notes in Computer Science, Klette, Reinhard and Peleg, Shmuel and Sommer, Gerald (Eds), pp. 201-209, 2001.

(Gerónimo et al., 2010) Gerónimo D., López A., Sappa A., and Graf T., "Survey of Pedestrian Detection for Advanced Driver Assistance Systems", Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.32, no.7, pp.1239-1258, 2010.

(Goldberg, 1989) Goldberg D., "Genetic Algorithms in Search, Optimization and Machine Learning", :Addison-Wesley, 1989.

(Goldberg et al., 2002) Goldberg S., Maimone M., and Matthies L., "Stereo vision and rover navigation software for planetary exploration", Aerospace Conference Proceedings, 2002. IEEE , vol.5, no., pp. 5-2025- 5-2036 vol.5, 2002

(Gong et al., 2007) Gong M., Yang R., Wang L., and Gong M., "A performance study on different cost aggregation approaches used in real-time stereo matching", IJCV, vol. 75, no. 2, pp. 283–296, 2007.

(Gong & Yang, 2005) Gong M., and Yang Y., "Near real-time reliable stereo matching using programmable graphics hardware", In: Proc. CVPR. pp. 924–931, 2005.

(Goulermas, 2000) Goulermas J., "Evolutionary techniques for the stereo-correspondence problem", Ph.D. dissertation, Control Syst.Centre, Univ. Manchester Inst. Sci. Technol., Manchester, U.K., 2000.

(Goulermas et al., 2005) Goulermas J., Liatsis P., and Fernando T. "A constrained nonlinear energy minimization framework for the regularization of the stereo correspondence problem", Circuits and Systems for Video Technology, IEEE Transactions on , vol.15, no.4, pp. 550- 565, 2005.

(Gu et al., 2008) Zheng Gu, Xianyu Su, Yuankun Liu, and Qican Zhang, “Local stereo matching with adaptive support-weight, rank transform and disparity calibration”, Pattern Recognition Letters, Vol. 29, No. 9, pp. 1230-12353, 2008.

(Guelch, 1991) Guelch E., “Results of test on image matching of ISPRS WG III/4”, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 46, pp. 1-18, 1991.

(Gupta & Cho, 2010a) Gupta R., and Cho S., “Real-time stereo matching using adaptive binary window”, 3D Data Processing, Visualization and Transmission, 2010.

(Gupta & Cho, 2010b) Gupta R., and Cho S, “A Correlation-Based Approach for Real-Time Stereo Matching”, Advances in Visual Computing, Vol 6454, LNCS, Bebis G. et al., Eds, Springer , pp. 129- 138, 2010.

H

(Haeusler & Klette, 2010) Haeusler R., and Klette R., “Benchmarking Stereo Data (Not the Matching Algorithms)”, Pattern Recognition, Vol. 6376, Lecture Notes in Computer Science, Goesele, Michael and Roth, Stefan and Kuijper, Ar and Schiele, Bernt and Schindler, Konrad Eds., Springer Berlin Heidelberg, pp. 383-392, 2010.

(Hanna,1974) Hanna M., “Computer Matching of Areas in Stereo Images”, PhD thesis, Stanford Univ., 1974.

(Harris & Stephens, 1988) Harris C., and Stephens M., “A combined corner and edge detector”, In Proceedings of The Fourth Alvey Vision Conference, pp. 147-151, Manchester, UK, 1988.

(Hartley, 1997) Hartley R., “In defence of the 8-point algorithm”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.19, pp. 580-593, 1997.

(Hartley & Zisserman, 2004) Hartley R., and Zisserman A., “Multiple view geometry in computer vision”, Second Edition, Cambridge University Press, 2004.

(Heikkilä, 2000) Heikkilä J, “Geometric Camera Calibration Using Circular Control Points”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, no. 10, pp. 1066-1077, 2000.

(Heitger et al., 1992) Heitger F., Rosenthaler L., von der Heydt R., Peterhans E., and Kuebler O. “Simulation of neural contour mechanism: From simple to end-stopped cells”, Vision Research, 32(5):963–981, 1992.

(Hirschmuller, 2001) Hirschmuller H., “Improvements in real-time correlation-based stereo vision”, Stereo and Multi-Baseline Vision – SMBV, Proceedings IEEE Workshop on , pp.141-148, 2001.

(Hirschmuller et al., 2002) Hirschmüller H., Innocent P., and Garibaldi J., “Real-Time Correlation-Based Stereo Vision with Reduced Border Errors”, Int. J. Comput. Vision 47, (1):229-24, 2002.

(Hirschmuller, 2003) Hirschmüller H., “Stereo Vision Based Mapping and Inmediate Virtual Walktrhoughs” PhD Thesis, De Monfort University, UK, 2003.

(Hirschmuller, 2005) Hirschmüller H., “Accurate and efficient stereo processing by semi-global matching and mutual information”, Computer Vision and Pattern Recognition, - CVPR. IEEE Computer Society Conference on , vol.2, no., pp. 807- 814 vol. 2, 2005.

(Hoang & McAllester, 2009.) Hoang T., and McAllester D., “Unsupervised Learning of Stereo Vision with Monocular Cues”, In A. Cavallaro, S. Prince and D. Alexander, editors, Proceedings of the British Machine Conference, BMVA Press, 2009.

(Horn et al.,1994) Horn J., Nafpliotis N., and Goldberg, D., “A niched Pareto genetic algorithm for multiobjective optimization”, Evolutionary Computation, 1994. IEEE

World Congress on Computational Intelligence., Proceedings of the First IEEE Conference on , vol. 1, pp.82-87, 1994.

(Hsieh et al., 1992) Hsieh Y., McKeown D., and Perlant F., “Performance evaluation of scene registration and stereo matching for cartographic feature extraction”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2):214–238, 1992.

(Horaud et al., 1990) Horaud R., Skordas T., and Veillon F., “Finding geometric and relational structures in an image”, In Proceedings of the 1st European Conference on Computer Vision, Antibes, France, pp. 374– 384, 1990.

(Hosni et al., 2009) Hosni A., Bleyer M., Gelautz M., and Rhemann C., “Local stereo matching using geodesic support weights”, Image Processing (ICIP), 16th IEEE International Conference on , pp. 2093-2096, 2009.

(Hosni et al., 2010) Hosni A., Bleyer M., and Gelautz M., “Near Real-Time Stereo With Adaptive Support Weight Approaches”, 3DVPT, 2010.

(Hosni et al., 2012) Hosni A., Rhemann C., Bleyer M., Rother C., and Gelautz, M., “Fast Cost-Volume Filtering for Visual Correspondence and Beyond”, Pattern Analysis and Machine Intelligence, IEEE Transactions on, pp. 3017- 3024, 2012.

(Howard et al., 2012) Howard T., Morfopoulos A., Morrison J., Kuwata Y., Villalpando C., Matthies L., and McHenry, M., “Enabling continuous planetary rover navigation through FPGA stereo and visual odometry”, Aerospace Conference, IEEE , pp.1-9, 2012.

(Humenberger et al., 2010) Humenberger M., Zinner C., Weber M., Kubinger W., and Vincze M., “A fast stereo matching algorithm suitable for embedded real-time systems”, Computer Vision and Image Understanding, vol. 114, no. 11, pp. 1180-1202, 2010.

I

(Intille & Bobick, 1994) Intille S., and Bobick A., “Disparity-space images and large occlusion stereo”, In J.-O. Eklundh, editor, European Conference on Computer Vision, Springer-Verlag, pp. 179-186, 1994.

(Isgro & Trucco 1999) Isgro F., and Trucco E., “Projective rectification without epipolar geometry”, Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. , vol.1, pp. 637-663, 1999.

(Isgro et al., 2001) Isgro F., Trucco E., and Li-qun Xu, “Towards teleconferencing by view synthesis and large-baseline stereo”, Image Analysis and Processing, Proceedings. 11th International Conference on , pp.198-203, 2001.

J

K

(Kanade & Okutomi,1994) Kanade T., and Okutomi M., “A stereo matching algorithm with an adaptive window: theory and experiment”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(9):920–932, 1994.

(Kang et al., 2008) Kang Y., Lee C., Ho, Y., “An Efficient Rectification Algorithm for Multi-View Images in Parallel Camera Array”, 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, pp.61-64, 28-30 2008.

(Keller et al., 2011) Keller C., Enzweiler M., Rohrbach M., Llorca D.F., Schnorr C., and Gavrila, D., “The Benefits of Dense Stereo for Pedestrian Detection”, Intelligent

Transportation Systems, IEEE Transactions on , vol.12, no.4, pp.1096-1106, . 2011

(Kelly, 2007) Kelly P., “Pedestrian detection and tracking using stereo vision techniques”, Ph.D. dissertation, Sch. of Electron. Eng., Dublin City Univ., Dublin, Ireland, 2007.

(Kelly et al., 2008) KellyP., O'Connor N., and Smeaton A., “A Framework for Evaluating Stereo-Based Pedestrian Detection Techniques”, Circuits and Systems for Video Technology, IEEE Transactions on , vol.18, no.8, pp.1163-1167, . 2008

(Kerr et al., 2008) Kerr D., Coleman S., and Scotney B., “Comparing Cornerness Measures for Interest Point Detection”, Machine Vision and Image Processing Conference, IMVIP, pp.105-110, 2008.

(Kitchen & Rosenfeld, 1982) Kitchen L., and Rosenfeld A., “Gray Level Corner Detection”, Pattern Recognition Letters, pp. 95-102, 1982.

(KITTI, 2012) URL: http://www.cvlibs.net/datasets/kitti/ , 2012.

(Klaus et al., 2006) Klaus A., Sormann M., and Karner K., “Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure”, Pattern Recognition - ICPR, 18th International Conference on, vol.3, no., pp.15-18, 2006.

(Klette et al., 2011) Klette R., Je Ahn, Haeusler R., Herman S., Jinsheng Huang, Khan W., Manoharan S., Morales S., Morris J., Nicolescu R., FeiXiang Ren; Schauwecker K., and Xi Yang; , “Advance in vision-based driver assistance”, Electric Technology and Civil Engineering (ICETCE), 2011 International Conference on , pp.987-990, 2011.

(Knowles & Corne 1999) Knowles J., and Corne D., “The Pareto archived evolution strategy: a new baseline algorithm for Pareto multiobjective optimisation”,

Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on , vol.1, pp. 98-105 vol. 1999.

(Koenderink & Doorn, 1987) Koenderink J., and van Doorn A., “Representation of Local Geometry in the Visual System”, Biological Cybernetics, vol. 55, pp. 367–375, 1987.

(Koninckx & Van Gool, 2006) Koninckx T., and Van Gool L., “Real-time range acquisition by adaptive structured light”, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.28, no.3, pp. 432-445, 2006.

(Kolmogorov &. Zabih, 2001) Kolmogorov V., and Zabih R., “Computing visual correspondence with occlusions using graph cuts”, In ICCV, pp. 508-515, 2001.

(Kolmogorov & Zabih, 2002) Kolmogorov V., and Zabih R., “Multi-camera scene reconstruction via graph cuts”, In Heyden A. et al., (Eds), ECCV, no. 2352 in LNCS, Springer, pp. 82-96, 2002.

(Kosov et al., 2009) Kosov S., Thormählen T., Seidel H., “Accurate Real-Time Disparity Estimation with Variational Methods”, Advances in Visual Computing Vol 5875 LNCS, Bebis G. et al. (Eds) Springer pp. 796-807, 2009.

(Kostliva et al., 2007) Kostliva J., Cech J., and Sara R., “Feasibility Boundary in Dense and Semi-Dense Stereo Matching”, Computer Vision and Pattern Recognition, - CVPR IEEE Conference on , pp.1-8, 2007.

(Kowalczuk et al., 2012) Kowalczuk J., Psota E., and Perez L., , “Real-time Stereo Matching on CUDA using an Iterative Refinement Method for Adaptive Support-Weight Correspondences”, Circuits and Systems for Video Technology, IEEE Transactions on , 2012.

L

(Ladicky et al., 2010) Ladicky L., Sturgess P., Russell C., Sengupta S., Bastanlar Y., Clocksin W., and Torr P., “Joint optimisation for object class segmentation and dense stereo reconstruction”. In BMVC, BMVA Press, 2010.

(Lan et al., 1995) Lan Z., Mohr R. and Remagnino P., “Robust Matching by Partial Correlation”, Proceedings of the sixth British Machine Vision Conference- BMVC, pp 651- 660, 1995.

(Leibe et al., 2007) Leibe B., Cornelis N., Cornelis K., and Gool L., “Dynamic 3D scene analysis from a moving vehicle”, In: Conference on Computer Vision and Pattern Recognition, 2007.

(Lindeberg, 1994) Lindeberg T., “Scale-space theory: A basic tool for analysing structures at different scales”, Journal of Applied Statistics, 21, 2 (1994), pp. 224–270.

(Little, 1992) Little J., “Accurate early detection of discontinuities”, Vision Interface, pp. 97–102, 1992.

(Liu & Klette, 2009) Liu Z. and Klette R., “Approximated Ground Truth for Stereo and Motion Analysis on Real-World Sequences”, Advances in Image and Video Technology, LNCS, vol. 5414, Wada T. et al., (Eds) Springer, pp. 874-885, 2009.

(Leclerc et al., 2000) Leclerc G., Quang-tuan Luong, Fua P., “Measuring the Self-Consistency of Stereo Algorithms”, European Conference on Computer Vision - ECCV, pp. 282-298, 2000.

(Leclercq et al., 2003) Leclercq P., and Morris J., “Robustness to noise of stereo matching”, Image Analysis and Processing, 2003.Proceedings. 12th International Conference on , pp. 606- 611, 2003.

(Lempitsky et al., 2007) Lempitsky V., Rother C., and Blake A., “LogCut - Efficient Graph Cut Optimization for Markov Random Fields”, Computer Vision – ICCV, IEEE 11th International Conference on, pp.1-8, 2007.

(Longuet-Higgins,1981) Longuet-Higgins H., “A computer program for reconstructing a scene from two projections”, Nature, vol. 293, pp. 133-135, 1981.

(Loop & Zhang, 1999) Loop C., and Zhengyou Zhang, “Computing rectifying homographies for stereo vision”, Computer Vision and Pattern Recognition - CVPR, IEEE Computer Society Conference on., pp. 125 – 131, 1999.

(López & Coello, 2009) López A., and Coello C, “ Study of preference relations in many-objective optimization”, In: Proc. Genetic and Evolutionary Computation Conference, pp. 611-618, 2009.

(Lowe, 1999) Lowe D., “Object recognition from local scale-invariant features”, In ICCV, pp. 1150-1157, 1999.

(Lowe, 2004) Lowe D., “Distinctive image features from scale invariant keypoints”, International Journal of Computer Vision, Vol.60, pp. 91–110, 2004.

(Luong & Faugeras, 1996) Luong Q., and Faugeras O., “The fundamental matrix: theory, algorithms, and stability analysis”, The International Journal of Computer Vision, vol.17, no.1, pp. 43-76, 1996.

M

(Maimone & Shafer, 1996) Maimone M., and Shafer S., “ECCV Workshop on Performance Characteristics of Vision Algorithms”, pp. 59 – 79, 1996.

(Malpica & Bovik, 2009) Malpica W., and Bovik A., “Range image quality assessment by Structural Similarity”, Acoustics, Speech and Signal Processing, ICASSP. IEEE International Conference on, pp.1149-1152, 2009.

(Mark et al., 1997) Mark W., McMillan L., and Bishop G., “Post-Rendering 3D Warping”, in Proc. of Symposium on Interactive 3D Graphics, pp. 7-16, 1997.

(Mark & Gavrila 2006) van der Mark W., and Gavrila D., “Real-time dense stereo for intelligent vehicles”, Intelligent Transportation Systems, IEEE Transactions on , vol.7, no.1, pp.38-50, 2006.

(Marr, 1982) Marr D., “Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information” W.H.Freeman & Co Ltd 1982.

(Marr & Hildreth, 1980) Marr D., and Hildreth E., “Theory of edge detection”, Proc. Royal Soc. London, B-207, pp.187-217, 1980.

(Matthies & Shafer, 1987) Matthies L., and Shafer S., “Error modeling in stereo navigation”, Robotics and Automation, IEEE Journal of , vol.3, no.3, pp.239-248, 1987.

(Matthies et al., 2008) Matthies L., Huertas A., Yang Cheng, and Johnson A., “Stereo vision and shadow analysis for landing hazard detection”, Robotics and Automation - ICRA. IEEE International Conference on, pp. 2735-2742, 2008.

(Meer & Georgescu, 2001) Meer P. and Georgescu B., "Edge detection with embedded confidence," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.23, no.12, pp. 1351-1365, 2001.

(Mei et al., 2011) Mei X., Sun X., Zhou M., Jiao S., Wang H., Zhang X., “On building an accurate stereo matching system on graphics hardware”, Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, pp. 467-474, 2011.

(Meesters et al., 2004) Meesters L., IJsselsteijn W., and Seuntiens P., “A survey of perceptual evaluations and requirements of three-dimensional TV”, *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.14, no.3, pp. 381-391, 2004.

(Mikolajczyk & Schmid, 2001) Mikolajczyk K., and Schmid C., “Indexing based on scale invariant interest points”, Computer Vision, - ICCV. Proceedings. Eighth IEEE International Conference on, vol.1, pp. 525-531, 2001.

(Mikolajczyk & Schmid, 2005) Mikolajczyk K., and Schmid C., “A performance evaluation of local descriptors”, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.27, no.10, pp.1615-1630, 2005.

(Min & Sohn, 2008) Min D., and Sohn K., “Cost aggregation and occlusion handling with wls in stereo matching”, Image Processing, IEEE Transactions on, vol. 17, no. 8, pp. 1431–1442, 2008.

(Mokhtarian & Suomela, 1998) Mokhtarian F., and Suomela R., “Robust image corner detection through curvature scale space”, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.20, no.12, pp.1376-1381, 1998.

(Mokhtarian & Mohanna, 2006) Mokhtarian F., and Mohanna F., “Performance evaluation of corner detectors using consistency and accuracy measures”, Computer Vision and Image Understanding, vol 102, no. 1, pp .81-94, 2006.

(Moravec, 1977) Moravec H., “Towards Automatic Visual Obstacle Avoidance”, Proc. Int’l Joint Conf. Artificial Intelligence, pp. 584, 1977.

(Moravec, 1979) Moravec H., “Visual Mapping by a Robot Rover”, International Joint Conference on Artificial Intelligence, pp. 598-600, 1979.

(Moravec, 1980) Moravec H., “Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover”, (Ph.D. thesis) Computer Science Department, Stanford University, 1980.

(Morales et al., 2009) Morales S., Vaudrey T., and Klette R., “Robustness evaluation of stereo algorithms on long stereo sequences”, Intelligent Vehicles Symposium, IEEE , pp.347-352, 2009.

(Morales & Klette, 2009) Morales S., and Klette R., “A Third Eye for Performance Evaluation in Stereo Sequence Analysis”, Computer Analysis of Images and Patterns vol 5702 LCNS, Springer Berlin Heidelberg, pp. 1078-1086, 2009.

(Morales & Klette, 2011) Morales S., and Klette R., “Ground Truth Evaluation of Stereo Algorithms for Real World Applications”, Computer Vision ACCV Workshop, vol 6469 LCNS Springer Berlin Heidelberg, pp. 152-162, 2011,

(Moreels & Perona, 2005) Moreels P., Perona P., “Evaluation of features detectors and descriptors based on 3D objects”, Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on , vol.1, pp. 800- 807, 2005.

(Mühlmann et al.,2002) Mühlmann K., Maier D., Hesser J., and Männer R., “Calculating dense disparity maps from color stereo images, an efficient implementation”, Int. J. Comput. Vis., vol. 47, no. 1, pp. 79–88, 2002.

(Mulligan et al., 2001) Mulligan J., Isler V., and Daniilidis K., “Performance evaluation of stereo for tele-presence”. In ICCV, vol. II, pp. 558–565, 2001.

(Mukherjee et al., 2010) Mukherjee D., Guanghui Wang, and Wu Q., “Stereo matching algorithm based on curvelet omposition and modified support weights”, Acoustics Speech and Signal Processing - ICASSP, IEEE International Conference on, pp.758-761, 2010.

N

(Nalpantidis & Gasteratos 2010a) Nalpantidis L., and Gasteratos A., “Biologically and psychophysically inspired adaptive support weights algorithm for stereo

correspondence", Robotics and Autonomous Systems, vol 58, no. 5, pp. 457-464 2010.

(Nalpantidis & Gasteratos 2010b) Nalpantidis L., and Gasteratos A., "Stereo vision for robotic applications in the presence of non-ideal lighting conditions", Image and Vision Computing, vol 28, no. 6, pp. 940-951, 2010.

(Nakamura et al., 1996) Nakamura Y., Matsuura T., Satoh K., and Ohta Y., "Occlusion detectable stereo - occlusion patterns in camera matrix", In CVPR, pp. 371–378, 1996.

(Nielsen et al., 2007) Nielsen M., Andersen H., Sl hter D., and Granum E., "Ground truth evaluation of computer vision based 3D reconstruction of synthesized and real plant images", Precision Agriculture, Kluwer Academic Publishers-Plenum Publishers, pp. 49-62, 2007

(Noble, 1988) Noble A., "Finding corners", Image Vision Comput 6(2):121- 128, 1988.

(Neilson & Yang, 2008) Neilson D., and Yee-Hong Yang, "Evaluation of constructable match cost measures for stereo correspondence using cluster ranking", Computer Vision and Pattern Recognition, - CVPR, IEEE Conference on , pp.1-8, 23-28, 2008.

O

(Olague & Trujillo, 2012) Olague G., and Trujillo L., "Interest point detection through multiobjective genetic programming", Applied Soft Computing, Vol. 12, No. 8, ust 2012, pp. 2566-2582.

(Okutomi & Kanade , 1993) Okutomi M., and Kanade T., “A Multiple Baseline Stereo”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 15, no. 4, pp. 353-363, 1993.

P

(Park & Han, 1998) Park J., and Han J., “Contour matching: a curvature-based approach”, Image and Vision Computing, Vol. 16, No. 3, pp. 181-189, 1998.

(Paris & Durand, 2009) Paris S., and Durand F., “A fast approximation of the bilateral filter using a signal processing approach”, International Journal Computer Vision (81): 24–52, 2009.

(Parreiras et al., 2006) Parreiras R., Maciel J., and Vasconcelos J., “The a posteriori ision in multiobjective optimization problems with smarts, promethee II, and a fuzzy algorithm”. In IEEE Trans. Magnetics, 42(4):1139-1142, 2006.

(Paris et al., 2008) Paris S., Kornprobst P., Tumblin J., and Durand F., “A gentle introduction to bilateral filtering and its applications”, In: SIGGRAPH Classes. Course material available online at http://people.csail.mit.edu/sparis/bf course. 2008

(Pfeiffer & Franke, 2010) Pfeiffer D., and Franke U., “Efficient representation of traffic scenes by means of dynamic Stixels”, In IEEE Intelligent Vehicles Symposium (IV), pp. 217–224, 2010.

(PtGrey, 2012) URL: http://www.ptgrey.com 2012

Q

(Qiang et al., 2011) Qiang Li, Biswas M., Pickering M., and Frater M., “Accurate depth estimation using structured light and passive stereo disparity estimation”, Image Processing - ICIP, 18th IEEE International Conference on , pp. 969-972, 2011

R

(Rachmawati & Srinivasan, 2006) Rachmawati L., and Srinivasan D, “Preference Incorporation in Multi-objective Evolutionary Algorithms: A Survey”, In: Proc. IEEE Congress On Evolutionary Computation, pp. 962--968 (2006)

(Ranftl et al., 2012) Ranftl R., Gehrig S., Pock T., and Bischof H., “Pushing the limits of stereo using variational stereo estimation”, IEEE Intelligent Vehicles Symposium (IV), pp. 401-407, 2012

(Rhemann et al., 2011) Rhemann C., Hosni A., Bleyer M., Rother C., and Gelautz, M., “Fast cost-volume filtering for visual correspondence and beyond”, Computer Vision and Pattern Recognition -CVPR, IEEE Conference on, pp. 3017-3024, 2011

(Richardt et al., 2010) Richardt C., Orr D., Davies I., Criminisi A., and Dodgson N. “Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid”, In Proceedings of the European Conference on Computer Vision- ECCV, 2010.

(Rodriguez & Aggarwal, 1990) Rodriguez J., and Aggarwal J., “Stochastic analysis of stereo quantization error”, Pattern Analysis and Machine Intelligence”, IEEE Transactions on , vol.12, no.5, pp.467-470, 1990.

(Roy & Cox, 1998) Roy S., and Cox I., “A maximum-flow formulation of the n-camera stereo correspondence problem”. In International Conference on Computer Vision - ICCV. pp. 492-499,1998.

(Rubner et al., 1998) Rubner Y., Tomasi C., and Guibas L., “A metric for distributions with applications to image databases”, Computer Vision, 1998. Sixth International Conference on, pp.59-66, 1998.

S

(Salvi et al., 2004) Salvi J., Pags J., and Batlle J., “Pattern codification strategies in structured light systems”, Pattern Recognit., vol. 37, no. 4, pp. 827–849, 2004.

(Salmen et al., 2009) Salmen J., Schlipsing M., Edelbrunner J., Hegemann S., and Lüke S., “Real-Time Stereo Vision: Making More Out of Dynamic Programming”, Computer Analysis of Images and Patterns, Vol 5702, LNCS, Jiang, Xiaoyi and Petkov, Nicolai (Eds.), Springer, pp. 1096-1103, 2009.

(Scharstein & Szeliski, 2002) Scharstein D., and Szeliski R., “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms”, Int’l Journal Computer Vision, vol. 47, no. 1, pp. 7-42, 2002.

(Scharstein & Szeliski, 2003) Scharstein D., and Szeliski R., “High-accuracy stereo depth maps using structured light”, Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on , vol.1, pp. 195-202, vol.1, 2003.

(Scharstein & Szeliski, 2012) Scharstein D., and Szeliski R. URL: http://vision.middlebury.edu/stereo 2012.

(Schmid & Mohr, 1997) Schmid C., and Mohr R., “Local gray value invariants for image retrieval”, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.19, no.5, pp.530-535, 1997.

(Schmid et al., 2000) Schmid C., Mohr R., and Bauckhage C., “Evaluation of Interest Point Detectors”, Int’l Journal Computer Vision, vol. 37, no. 2, pp. 151-172, 2000.

(Schreer et al., 2006) Schreer O., Fehn C., Atzpadin N., Muller M., Smolic A., Tanger R., and Kauff P., “A Flexible 3D TV System for Different Multi-Baseline Geometries”, Multimedia and Expo, IEEE International Conference on , pp.1877-1880, 2006.

(Schneider et al., 2011) Schneider N., Gehrig S., Pfeiffer D., and Banitsas K., “An Evaluation Framework for Stereo-Based Driver Assistance”, Theoretical Foundations of Computer Vision, vol. 7474 of Lecture Notes in Computer Science, pp. 27-51. Springer, 2011.

(Sellent & Wingbermüle, 2012) Sellent A. and Wingbermüle J., “Quality Assessment of Non-dense Image Correspondences”, In Procs. ECCV, LNCS, Fusiello A. et al., (Eds), Springer, vol. 7584, , pp. 114-123, 2012.

(Shah & Jain, 1984) Shah M., and Jain R., “Detecting time-varying corners”, Comput. Vis. Graph. Image Process 28:345-355 1984

(Shen et al., 2011) Shen Y., Chaohui L., Xu P., and Xu L., “Objective Quality Assessment of Noised Stereoscopic Image”, In: Proc. Third Intl Conf. on Measuring Technology and Mechatronics Automation, pp. 745–747, 2011.

(Smith & Brady, 1997) Smith S., and Brady J., “SUSAN - A new approach to low level image processing”, International Journal of Computer Vision 23 (1):45–78, 1997.

(Smith et al., 2009) Smith M., Baldwin I. Churchill W., Paul R.,. and Newman P., “The new college vision and laser data set”, The International Journal of Robotics Research, Vol 28 No 5, SAGE Publications, Inc. , pp. 595-599, 2009.

(Sun et al., 2005) SunJ., Yin Li, Kang S., and Heung-Yeung Shum, “Symmetric stereo matching for occlusion handling”, Computer Vision and Pattern Recognition, - CVPR. IEEE Computer Society Conference o , vol.2, pp. 399- 406, 2005.

(Steingrube et al., 2009) Steingrube P., Gehrig S.. and Franke U., “Performance Evaluation of Stereo Algorithms for Automotive Applications”, Computer Vision Systems, Vol. 5815, LNCS, Fritz et al. (Eds), Springer, pp. 285-294, 2009.

(Szeliski, 1999) Szeliski R., “Prediction error as a quality metric for motion and stereo”, in Proc. Seventh International Conference on Computer Vision - ICCV, pp. 781-788, 1999.

(Szeliski & Zabih, 1999) Szeliski R., and Zabih R., “An experimental comparison of stereo algorithms”, In International Workshop on Vision Algorithms, pp. 1–19, Springer, 1999.

(Szeliski, 2010) Szeliski R., “Computer Vision: Algorithms and Applications”, Springer, 2010.

T

(Taguchi et al., 2008) Taguchi, Y., Wilburn, B., and Zitnick, C., “Stereo reconstruction with mixed pixels using adaptive over-segmentation”, Computer Vision and Pattern Recognition - CVPR. IEEE Conference on, pp.1-8, 2008.

(Tao and & Sawhney, 2000) Tao H., and Sawhney H., “Global matching criterion and color segmentation based stereo”, Proc. Workshop on the Application of Computer Vision, pp. 246-253, 2000.

(Tao et al., 2011) Tao N., Hongyan Zhang, Songyue Liu, and Yamada H., “Teleoperation system with virtual reality based on stereo vision”, Transportation, Mechanical,

and Electrical Engineering - TMEE), International Conference on, pp.494-497, 2011.

(Taylor & Bhusnurmath, 2008) Taylor C., and Bhusnurmath A., “Solving Image Registration Problems Using Interior Point Methods”, LNCS, Forsyth D., et al. (Eds), Spinger, pp.638-651, 2008.

(Terrile & Noraky, 2012) Terrile, R., and Noraky, J., “Immersive telepresence as an alternative for human exploration”, Aerospace Conference, IEEE , pp.1-11, 2012.

(Tian & Huhns, 1986) Tian Q., and Huhns M., “Algorithms for subpixel registration”, CVGIP, 35:220–233, 1986.

(Tissainayagam & Suter, 2004) Tissainayagam P., and Suter D., “Assessing the performance of corner detectors for point feature tracking applications”, Image and Vision Computing, Elsevier, Vol. 22, pp. 663-679, 2004.

(Thacker et al., 2008) Thacker N., Clark A., Barron J., Beveridge J., Courtney P., Crum W., Ramesh V., and Clark C, “ Performance characterization in computer vision: A guide to best practices”, Computer Vision Image Understanding Vol. 109, pp. 305–334, 2008.

(Tomasi & Manduchi, 1998) Tomasi C., and Manduchi R., “Bilateral filtering for gray and color images”. In: Proc. International Conference Computer Vision - ICCV, pp. 839–846, 1998.

(Tomasi & Manduchi, 1999) Manduchi R., and Tomasi C., “Distinctiveness maps for image matching”, In International Conference on Image Analysis and Processing, pp. 26–31, 1999.

(Tomasi , T. Kanade) Tomasi C., and Kanade T., “Detection and Tracking of Point Features”, Carnegie Mellon University, Tech. Report CMU-CS-91-132 – Unpublished, 1991.

(Tombari el al., 2007) Tombari F., Mattoccia S., and Di Stefano L., “Segmentation-based adaptive support for accurate stereo correspondence”, in PSIVT, pp. 427,438 2007.

(Tombari et al., 2008) Tombari F., Mattoccia S., Di Stefano L., and Addimanda E., “Near real-time stereo based on effective cost aggregation”, Pattern Recognition, 2008. ICPR, 19th International Conference on, pp.1-4, 2008.

(Torr &. Murray,1997) Torr P., and Murray D., “The development and comparison of robust methods for estimating the fundamental matrix”, International Journal of Computer Vision, vol. 24, pp. 271-300, 1997.

(Trajković & Hedley, 1998) Trajković M., and Hedley M., “Fast corner detection”, Image and Vision Computing, Vol. 16, No. 2, pp. 75-87, 1998.

(Trucco et al., 2013) Trucco E., Ruggeri A., Karnowski T., Giancarlo L., Chaum E., Hubschman J. P., Al-Diri B., Cheung C. Y., Wong D., Abramoff M., Lim G., Kumar D., Burlina P., Bressler N. M., Jelinek H., Maiaudeau F., Quellec G., MacGillivray T. J., and Dhillon B., “Validating Retinal Fundus Image Analysis Algorithms: Issues and a Proposal”, Investigative Ophthalmology & Visual Science, vol. 54, pp. 3546-3559, 2013.

(Trucco & Verri 1998) Trucco E., and Verri A., “Introductory Techniques for 3-D Computer Vision” Prentice Hall, 1998.

(Trujillo & Izquierdo, 2005) Trujillo M., and Izquierdo E., “Combining K-means and semivariogram-based grid clustering”, ELMAR, 47th International Symposium, pp.9-12, 2005.

U

V

(Vaudrey et al., 2008) Vaudrey T., Rabe C., Klette R., and Milburn, J., “Differences between stereo and motion behaviour on synthetic and real-world stereo sequences”, Image and Vision Computing New Zealand, IVCNZ. 23rd International Conference, pp.1-6, 2008.

(Veldhuizen & Lamont, 2000) Van Veldhuizen D., and Lamont G., “On measuring multiobjective evolutionary algorithm performance”, Evolutionary Computation Proceedings of the Congress on , vol.1, pp. 204-211, 2000.

(Veldhuizen et al., 2003) Van Veldhuizen, D., Zydallis J., and Lamont G., “Considerations in engineering parallel multiobjective evolutionary algorithms”, Evolutionary Computation, IEEE Transactions on , vol.7, no.2, pp. 144- 173, 2003.

(Veksler, 2002) Veksler O., “Stereo correspondence with compact windows via minimum ratio cycle”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 12, pp. 1654–1660, 2002.

(Viola & Wells, 1995) Viola P. and Wells W., “Alignment by maximization of mutual information”, In Computer Vision Proceedings., Fifth International Conference on, pp.16-23, 1995.

W

(Wang, 2004) Wang K., “Adaptive stereo matching algorithm based on edge detection”. International Conference on Image Processing - ICIP Vol 2, pp. 1345–1348, 2004.

(Wang & Brady, 1994) Wang H., and Brady M., “A practical solution to corner detection”, IEEE International Conference Image Processing- ICIP., vol.1, pp.919-923 vol.1, 1994.

(Wang & Brady, 1995) Wang H., and Brady M., “Real-time corner detection algorithm for motion estimation”, Image and Vision Computing, Vol. 13, No. 9, 1995.

(Wang & Bovik , 2002) Wang Z., and Bovik A., “A universal image quality index”, IEEE Signal Process. Lett., vol. 9, no. 3, pp.81-84, 2002.

(Wang et al., 2002) Wang Z.; Bovik A., and Lu L., “Why is image quality assessment so difficult?”, Acoustics, Speech, and Signal Processing (ICASSP), IEEE International Conference on , vol.4, no., pp.3313-3316, 2002.

(Wang et al., 2003) Wang Z., Simoncelli E., and Bovik A., “Multiscale structural similarity for image quality assessment”, Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on , vol.2, pp. 1398- 1402 Vol.2, 2003

(Wang et al., 2004) Wang Z., Bovik A., Sheikn H., and Simocell E., “ Image Quality Assessment: From Error visibility to Structural Similarity”. IEEE Trans. on Image Processing 13(4):600–612, 2004.

(Wang et al., 2006a) Wang L., Gong M., Gong M., Yang R., “How Far Can We Go with Local Optimization in Real-Time Stereo Matching”, 3D Data Processing, Visualization, and Transmission, Third International Symposium on , pp.129-136, 14-16 e 2006

(Wang et al., 2006b) Wang L., Liao M., Gong M., Yang R., and Nistér D., “High-quality real-time stereo using adaptive cost aggregation and dynamic programming”, In: Proc. 3DPVT, pp. 798–805, 2006.

(Wang & Bovik, 2009) Wang Z., and Bovik A., “Mean squared error: love it or leave it? - A new look at signal fidelity measures”, IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, 2009.

(Wang & Li, 2011) Wang Z., and Li Q., “Information Content Weighting for Perceptual Image Quality Assessment”, Image Processing, IEEE Transactions on, On page(s): 1185 - 1198 Vol. 20, no. 5, 2011.

(Woodward et al., 2006) Woodward A., Leclercq P., Delmas P., and Gimel'farb, G., “Generation of An Accurate Facialground Truth for Stereo Algorithm Evaluation”, Computer Vision and Graphics, Vol 32, Computational Imaging and Vision, Springer, pp. 534-539, 2006 .

X

(Xu & Zhang, 1996) Xu G., and Zhang Z., “Epipolar Geometry in Stereo, Motion, and Object Recognition: A Unified Approach”, Kluwer Academic Publishers Norwell, USA, 1996.

(Xu et al., 2002) Xu Y., Wang D., Feng T., and Shum H., “Stereo Computation using Radial Adaptive Windows”, Proc. Int’l Conf. Pattern Recognition, vol. 3, pp. 595-598, 2002.

Y

(Yamaguchi et al., 2012) Yamaguchi K., Hazan T., McAllester D., and Urtasun, R., “Continuous Markov Random Fields for Robust Stereo Estimation”, In Proc. ECCV, Vol 7576, LNCS, pp. 45-58, 2012.

(Yang et al., 2006) Yang Q., Wang L., and Yang R., “Real-time Global Stereo Matching Using Hierarchical Belief Propagation”. In Mike Chantler, Bob Fisher and Manuel Trucco, editors,Proceedings of the British Machine Conference, pp. 101.1-101.10, BMVA Press, 2006.

(Yang et al., 2007) Yang Q., Yang R., Davis J., and Nistér D., “Spatial-depth super resolution for range images”. Computer Vision and Pattern Recognition, - CVPR, IEEE Conference on, pp.1-8, 2007.

(Yang et al., 2008) Yang Q., Engels C., and Akbarzadeh A., “Near Real-time Stereo for Weakly-Textured Scenes”. In M. Everingham and C. Needham, editors, Proceedings of the British Machine Conference - BMVC, pp. 72.1-72.10. BMVA Press, 2008.

(Yang et al., 2009) Yang Q., Liang Wang, Ruigang Yang, Stewenius H., Nister D., “Stereo Matching with Color-Weighted Correlation, Hierarchical Belief Propagation, and Occlusion Handling”, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.31, no.3, pp.492-504, 2009.

(Yoon & Kweon, 2005) Yoon K., and Kweon I., “Locally adaptive support-weight approach for visual correspondence search”, Proc. CVPR, pp. 924-931, 2005.

(Yoon & Kweon, 2007) Yoon K., and In So Kweon, “Stereo Matching with the Distinctive Similarity Measure”, Computer Vision, ICCV. IEEE 11th International Conference on, pp.1-7, 2007.

(Yu et al., 2010) Yu W., Tsuhan Chen, Franchetti F., Hoe J., “High Performance Stereo Vision Designed for Massively Data Parallel Platforms”, Circuits and Systems for Video Technology, IEEE Transactions on , vol. 20, no.11, pp.1509-1519, 2010.

Z

(Zabih & Woodfill, 1994) Zabih R., and Woodfill J., “Non-parametric Local Transforms for Computing Visual Correspondence”, European Conference on Computer Vision, pp. 151-158, 1994.

(Zhang,1998) Zhang Z., “Determining the epipolar geometry and its Uncertainty: a review”, International Journal of Computer Vision, vol.27, pp. 161-195, 1998.

(Zhang et al.,1995) Zhang Z., Deriche R., F eras O., and Luong Q., “A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry”, Artificial Intelligence, vol.78, pp.87-119, 1995.

(Zhang, 2000) Zhang Z., “A flexible new technique for camera calibration”, IEEE Trans. Pattern Anal. Mach. Intell, vol. 22, no. 11, pp. 1330-1334, 2000.

(Zhang, 2004) Zhang Z., “Camera calibration with one-dimensional objects”, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.26, no.7, pp.892-899, 2004.

(Zhang et al., 2009a) Zhang K., Jiangbo Lu, and Lafruit G., “Cross-Based Local Stereo Matching Using Orthogonal Integral Images”, Circuits and Systems for Video Technology, IEEE Transactions on , vol.19, no.7, pp.1073-1079, 2009.

(Zhang et al., 2009b) Zhang K., Jiangbo Lu, Lafruit, G., Lauwereins R., and Van Gool L., “Real-time accurate stereo with bitwise fast voting on CUDA”, Computer Vision Workshops (ICCV Workshops), IEEE 12th International Conference on, pp.794-800, 2009.

(Zhang et al., 2009c) Zhang Z., Hou C., Shen L., and Yang J., "An Objective Evaluation for Disparity Map Based on the Disparity Gradient and Disparity Acceleration," Information Technology and Computer Science – ITCS, International Conference on , vol.1, pp.452-455, 2009.

(Zhao & Thorpe, 2000) Zhao L., and Thorpe C., “Stereo and neural network-based pedestrian detection”, IEEE Trans. Intell. Transportation Syst., vol. 1, no. 3, pp. 148–154, 2000.

(Zou et al., 2008) Zou L., Jie Chen, Juan Zhang, Li-hua Dou, “The Comparison of Two Typical Corner Detection Algorithms”, Intelligent Information Technology Application, 2008. IITA '08. Second International Symposium on , vol.2, pp.211-215, 2008.

(Zuniga & Haralick, 1983) Zuniga O., and Haralick R., “Corner detection using the facet model”, Proc. Conf. Part. Recog. Image Process. pp. 30-37, 1983.